

# Hyperspectral Cytometry

Gérald Grégori, Bartek Rajwa, Valery Patsekin, James Jones,  
Motohiro Furuki, Masanobu Yamamoto and J. Paul Robinson

**Abstract** Hyperspectral cytometry is an emerging technology for single-cell analysis that combines ultrafast optical spectroscopy and flow cytometry. Spectral cytometry systems utilize diffraction gratings or prism-based monochromators to disperse fluorescence signals from multiple labels (organic dyes, nanoparticles, or fluorescent proteins) present in each analyzed bioparticle onto linear detector arrays such as multianode photomultipliers or charge-coupled device sensors. The resultant data, consisting of a series of spectral fingerprints characterizing every analyzed cell, are not compensated by employing the traditional cytometry approach, but rather are spectrally unmixed utilizing algorithms such as constrained Poisson regression or non-negative matrix factorization. Although

---

G. Grégori · V. Patsekin · J. Paul Robinson  
Department of Basic Medical Sciences, College of Veterinary Medicine,  
Purdue University, West Lafayette, IN 47907, USA

J. Jones · J. Paul Robinson  
Weldon School of Biomedical Engineering, Purdue University, West Lafayette,  
IN 47907, USA

B. Rajwa  
Bindley Bioscience Center, Purdue University, West Lafayette, IN 47907, USA

G. Grégori  
Mediterranean Institute of Oceanography, Aix-Marseille Université,  
Université du Sud-Toulon Var, CNRS, IRD, Marseille 13288, France

M. Furuki  
Life Science Business Div, MBU, Sony Corporation, Tokyo 141-0001, Japan

M. Yamamoto  
Phototek Laboratory Inc, Yokohama 220-0003, Japan

J. Paul Robinson (✉)  
Purdue University Cytometry Laboratories, 1300 Lynn Hall, West Lafayette,  
IN 47907-2057, USA  
e-mail: jpr@flowcyt.cyto.purdue.edu

implementations of spectral cytometry were envisioned as early as the 1980s, only recently has the development of highly sensitive photomultiplier tube arrays led to design and construction of functional prototypes and subsequently to introduction of commercially available systems. This chapter summarizes the historical efforts and work in the field of spectral cytometry performed at Purdue University Cytometry Laboratories and describes the technology developed by Sony Corporation that resulted in release of the first commercial spectral cytometry system—the Sony SP6800. A brief introduction to spectral data analysis is also provided, with emphasis on the differences between traditional polychromatic and spectral cytometry approaches.

## Contents

1	Spectral Analysis by Cytometry .....	192
2	Modern Hyperspectral Cytometry .....	194
3	Practical Issues of Hyperspectral Cytometry .....	196
4	Basic Analytical Problems to be Addressed .....	197
5	Hyperspectral Cytometry Instrumentation and Data Acquisition .....	198
5.1	PUCL Prototype .....	198
5.2	Sony SP6800 Spectral Analyzer .....	201
6	Spectral Data Analysis .....	203
7	Potential Applications .....	208
	References .....	208

## 1 Spectral Analysis by Cytometry

Analytical flow cytometry (FC) is one of the most powerful single-cell analysis techniques available. Researchers in the life sciences, including basic cell biology, molecular biology and genetics, immunology, plant science, microbiology, environmental sciences, and oceanography, depend on FC to quantify cellular phenotypes and physiological responses of individual cells (Shapiro 2003). In order to perform quantitative measurements, FC relies on optical properties of biological particles (such as bacteria, algae, or mammalian cells) suspended in a liquid medium. On most currently available instruments, the forward-angle light scatter provides a proxy for the size of the particles, and the side-angle light scatter is used to characterize particle shape and internal structure. Both signals are created by interactions between individual bioparticles and a light beam produced by an external light source. Various fluorescence emission signals can be simultaneously collected following this excitation.

As the bioparticles flow rapidly through the instrument detection chamber (often called a flow cell) at rates of up to several thousand particles per second, the fluorescence emission data are automatically collected using photomultiplier tube (PMT) arrays, quantified and digitized via fast electronics, and eventually stored

on a computer for further statistical analysis. These data are subsequently processed by an operator using one of several dedicated software packages capable of visualizing and discriminating various populations of particles according to their optical properties. Cells or particles can be classified based on morphology, abundance of fluorescence labels, physiology, functional activity, or expression of certain cell-surface or internal antigenic determinants. The overall aim of cytometry analysis is to characterize heterogeneous cellular populations by decomposing them into a set of phenotypically different, but internally similar, groups described by biological function.

Although FC is perhaps the most widely used method for phenotypic analysis and classification, the core technology has remained essentially unchanged until recently. In almost all the current commercial FC systems, scatter and fluorescence signals travel down an optical pathway through a set of dichroic filters, each of which splits the incoming signal into two directions according to the wavelength bands selected. A signal from each fluorochrome is redirected in this manner until it reaches a dedicated point of acquisition where it is filtered through a specific band-pass filter of a desired wavelength before being collected on a dedicated photodetector that provides a current output proportional to light intensity. The charge or the current signal is subsequently converted into a voltage that can be readily digitized by an analog-to-digital converter and finally is recorded by a computer. In most commercial instruments, photodiodes detect the bright forward light-scatter signals, and photomultiplier tubes are used to collect the weak fluorescence emission signals; the latter typically require amplification (alternative single-cell technologies that are not dependent upon fluorescence, such as mass cytometry, are discussed elsewhere in this volume).

Over the past three decades, FC technology has developed from single-color (single band, single-fluorescence intensity) measurement systems through two-, three-, and four-color instruments to the newest benchtop instruments with 10 to 12 fluorescence detection channels. Although a collection of as many as 17 simultaneous separate fluorescence signals has been reported in a traditional cytometry experiment, typical FC systems collect a more manageable number of bands, typically between 5 to 10 (Perfetto et al. 2004; Roederer et al. 1997; De Rosa et al. 2001; Wang et al. 2009).

Despite the enormous progress in multiband (also called “polychromatic”) cytometry, it has been recognized for many years that collection of the full emission spectra would provide significantly more information than measurement of just few predefined bands. This type of collection would also allow for more flexible instrument design. In 1979, Wade et al. reported the fluorescence spectrum recorded for particles in a flow system (Wade et al. 1979); however, the instrumentation recorded only integrated spectra from a large number of particles. Since data collection was not achieved at the individual particle level, the sample was obviously assumed to be homogeneous. In 1986, Steen and Stokke were able to measure averaged fluorescence spectra of rat thymocytes. They used a custom-built cytometer equipped with a grating monochromator (Steen 1986; Stokke and Steen 1986). In 1990, Buican proposed the use of a Fourier-transform interferometer to

collect single-cell spectra, but in practical performance, the design had severe limitations as the cells had to remain in the laser beam for a relatively long time in order to obtain a measurable signal set (Buican 1990). By comparison, the time period available for the laser excitation on current high-speed FC systems lasts only from 1 to 10 microseconds. In 1996, Gauci et al. described a system based on a flint-glass prism and an intensified photodiode array (Gauci et al. 1996). Again, the low data acquisition rate precluded practical use, and, in addition, the efficiency of the photodiodes was inadequate. The same year Asbury et al. reported measurement of spectra of cells and chromosomes using a monochromator (Asbury et al. 1996). However, the design required that the wavelength be changed during the course of measurement, making continuous flow measurements impossible (Asbury et al. 1996; Gauci et al. 1996). The technique was limited to measurement of just a single band of fluorescence from any single particle. Other groups, including those from SoftRay Inc. and the Universities of Wyoming and Utah, pursued another prism-based concept, but no subsequent data were reported (Johnson et al. 2001).

## 2 Modern Hyperspectral Cytometry

The introduction of multianode photomultipliers by Hamamatsu (H7260 series) reinvigorated the attempts to construct a hyperspectral flow cytometer able to collect an approximation of a full spectrum from every single bioparticle in flow. In the early 2000s, researchers at Purdue University Cytometry Laboratories began work on hardware and software prototypes for fast classification of hydrodynamically focused bioparticles using a spectral detector extension attached to a commercial FC system (an EPICS Elite cell sorter, from Beckman Coulter). The concept assumed utilization of the recently available first-generation, 32-channel multianode PMT (Hamamatsu) that had also been used in the field of confocal microscopy (Robinson et al. 2007). The design rationale was to reduce the complexity of FC optical pathways by reducing the number of elements and replacing them with a single multiband detector capable of providing sufficient sensitivity, portability, and robustness (Grégori et al. 2012). The preliminary results documenting the early work on the multispectral FC were presented in 2004.<sup>1</sup> The data demonstrated that the technology had potential for future commercialization (Robinson 2004). As second-generation multianode PMTs offered better sensitivity, the second prototype was capable of simultaneously collecting 32 bands of fluorescence from each flowing particle in less than 5  $\mu$ s (Robinson et al. 2005). This created new opportunities for analysis and characterization of cells in a high-throughput and high-content setting, but it also required advanced control software capable of handling increased numbers of parameters.

---

<sup>1</sup> Presented at the International Congress of Analytical Cytology, May 23–28, 2004, Montpellier, France.

The work on spectral cytometry was also progressing in other laboratories. In 2006, Goddard et al. presented an alternative design concept employing a diffraction grating and a charge-coupled device (CCD) detector. This device dispersed the collected signals (fluorescence and side-scattered light) onto a CCD image sensor coupled to a spectrograph (Goddard et al. 2006). The design of the instrumentation involved minimal modifications around the flow chamber and collection optics of a conventional flow cytometer. Unfortunately, the flow rate was highly restricted owing to limited sensitivity of the CCD. Recently, John Nolan and collaborators demonstrated a new-generation, CCD-based spectral cytometry system. In this implementation, a broadband volume-phase holographic grating is interfaced with an electron-multiplying CCD detector. The system offers spectral resolution of approximately 11 nm (Nolan et al. 2013).

Although most of the efforts in spectral cytometry development remained focused on fluorescence, in 2008 John Nolan's group at La Jolla Bioengineering Institute presented a Raman spectral flow cytometry concept (RSFC) in which surface-enhanced Raman detection (SERS) and flow cytometry were combined (Watson et al. 2008). The CCD-based detector on the system was sensitive enough to detect SERS spectra in samples containing nanoparticle tags bound to microbeads. It was also capable of measuring Raman spectra from particles bearing as few as 200 Raman tags and had an integration time as short as 100 microseconds. Results obtained with the instrument indicated that it could detect more probes in the spectral range used than traditional fluorescence-based systems, thus offering a powerful tool for signal multiplexing. The development of robust tags remains an important challenge, however, as nanoparticle-based SERS labels tend to be relatively heterogeneous compared to organic fluorochromes or fluorescent proteins, which can be prepared with higher purity. Even though researchers have made SERS systems reproducible, the processes still require significant development (Brown and Doorn 2008a, b).

Although the well-established cytometry vendors largely ignored the new technology, Sony Corp.—a relative new comer in the field of cytometry—pursued an advanced hyperspectral design of their own. The Sony concept utilized a multianode PMT and featured a very complex prism-based monochromator. Sony demonstrated a prototype instrument and reported on hyperspectral technology during the ISAC congress in Seattle in 2010<sup>2</sup> and announced the launch of the new hyperspectral flow cytometer product—a SP6800 Spectral Cell Analyzer—in 2012. Although most of the mentioned implementations provided valuable scientific contributions, only the design using multiarray PMT technology has impacted the development of new generations of commercial instruments, as exemplified by the multispectral cytometry system designed by Sony.

---

<sup>2</sup> Motohiro Furuki, Shingo Imanishi, Masaya Kakuta, Masanobu Yamamoto, Yohei Morita, Yuji Yamazaki, Yumiko Ishii, Hiromitsu Nakauchi, "Hyper-Spectral Flow Cytometer with a Microfluidic Chip." Presented at CYTO2010, XXV Congress of International Society for Advancement of Cytometry, May 9–12, 2010, Seattle, WA, USA.

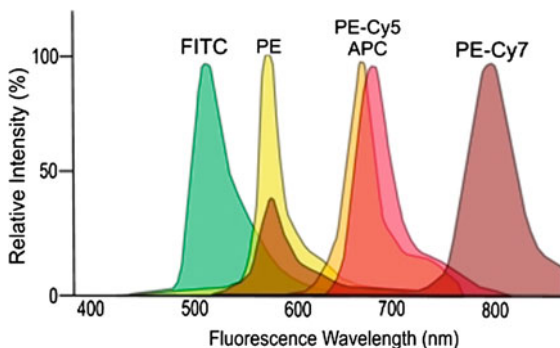
### 3 Practical Issues of Hyperspectral Cytometry

The technical aspects of hyperspectral flow cytometry are often discussed and compared with analogous spectral imaging techniques. Indeed, in the fields of high-resolution optical microscopy and small-animal imaging, one of the most significant recent technical developments has been the commercialization and wide acceptance of spectral imaging approaches (Zimmermann et al. 2003). However, the fact remains that the tools and techniques developed in spectral imaging are not readily transferable to the realm of FC owing to the following issues:

- In flow cytometry, the time for data collection is in the microsecond range. The particles or cells pass through a liquid-handling system and hydrodynamic forces within the flow chamber result in a single, central core stream. Once this hydrodynamic focusing has been accomplished, the particles (cells) usually pass very quickly (within a few microseconds) through a very narrow and focused beam of intense laser light, during which time a large number of variables—such as light scatter and spectral signatures—are collected and recorded. This high speed of flow is the core feature of cytometry that allows for the collection of data on several thousand particles per second.
- Separation of all the optical signals must be achieved within the time scale of the measurement system (i.e., a few microseconds), which eliminates all the tunable filter approaches.
- Significantly, many more fluorescence labels are used simultaneously in flow cytometry experiments than in typical imaging-based cytometry measurements.
- Cells cannot be analyzed more than once in regular FC systems (i.e., one cannot average multiple measurements from a single cell). Thus, one must collect all required data in a single measurement cycle. In imaging systems it is possible to scan, rescan, and average up to the photobleaching limits.
- In flow cytometry, every particle (cell) is a distinctive entity; FC allows true single-cell analysis. Each signal from every cell in a population is considered unique; therefore, it is not acceptable to average signals from multiple cells, and it is not possible to achieve the same population classification results if one does so.
- Cytometry is inherently quantitative. Therefore reproducibility and standardization issues are fundamental. On the other hand, spectral analysis methods in imaging are often used to provide qualitative results.

Most previous attempts at implementing spectral detection in flow cytometry failed to result in an instrument that could perform multicolor measurements with the required sensitivity and speed. The family of Hamamatsu's multiarray PMTs used in the prototype developed at Purdue as well as in a commercial

**Fig. 1** Typical extent of overlap of fluorescence for various molecules currently used in flow cytometry



implementation build by Sony (in a heavily modified version) are the first photodetectors that could be employed in multispectral cytometry setting.

#### 4 Basic Analytical Problems to be Addressed

Both the absorption and emission spectra of the fluorochromes used in FC may carry valuable spectral information about tagged biological particles. The commonly used optical design of FC instruments requires that researchers employ a series of fluorochromes that have narrow excitation maxima and produce reasonably narrow emission bands within the sensitivity of the detector.

To achieve multiplexing and perform experiments with several fluorescent probes, a variety of excitation sources (multiple lasers offering multiple wavelengths) and a well-designed panel of fluorescence probes with minimally overlapping emission spectra must be used. When combining several labels in a sample, however, overlap of the emissions always occurs to some extent (Fig. 1). Indeed, most currently used fluorescent probes are organic and tend to have rather broad excitation and emission spectra. Therefore, it is virtually impossible to measure a signal from one fluorochrome while completely excluding the emission from all the others. The experimental setup, as well as the proper linear unmixing of the signals (compensation), can be very difficult, as exemplified by the complicated configurations of systems employed to perform highly polychromatic cytometry analysis (Perfetto et al. 2004; Roederer 2001; Roederer et al. 1997; De Rosa et al. 2001). As researchers attempt to use more fluorochromes simultaneously, current techniques utilizing multiple individual PMTs are reaching their limit and are becoming extremely complex, expensive, and difficult to scale further. To harness the full potential of new probes, such as nanocrystals, and to perform multiparametric experiments easily and effectively, alternative technologies must be considered (Bernstein and Hyun 2012; Nolan et al. 2013).

## 5 Hyperspectral Cytometry Instrumentation and Data Acquisition

### 5.1 PUCL Prototype

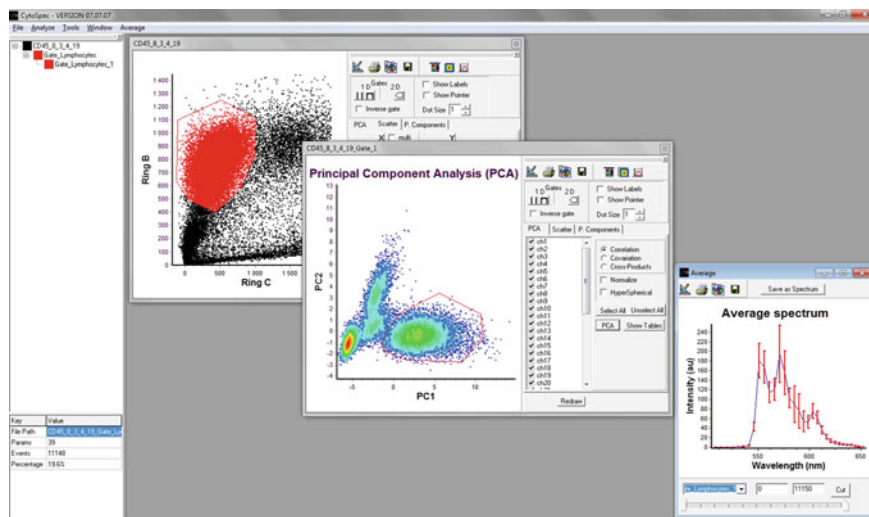
The first prototype of a hyperspectral flow cytometer developed at Purdue University Cytometry Laboratories (PUCL) utilized a heavily modified EPICS Elite cell sorter (Beckman Coulter). The data collection unit was comprised of a traditional custom-built polychromatic detection system using dichroic filters and a hyperspectral subunit designed at PUCL. The traditional detection module was based on a 30-nm FWHM polychromator from Asahi Spectra USA, Inc. This unit was equipped with six photodetectors (PMTs) and a set of band-pass filters (525/30 nm, optimized for FITC; 575/30 nm, PE; 620/30 nm, PE-Texas Red; 675/30 nm, PE-Cy5; 767/30 nm, PE-Cy7). To split the fluorescent signal coming from the measured particles in order to perform simultaneous measurement using the two units (the 32-channel PMT and the six-channel detection module), a 50/50 beam splitter was placed between the 32-channel spectral subsystem and the six-channel device. The EPICS Elite cell sorter was interfaced with these two devices to allow for simultaneous spectral and polychromatic data collection. Equivalent cell classification results were obtained using the significantly simplified spectral optical path and traditional detection components.

The spectral subsystem was comprised of a phase-volume holographic grating (Kaiser Optical Systems), which dispersed the signal onto a Hamamatsu 7260-01 32-channel multianode PMT. This multianode linear array offered a cathode sensitivity of 250  $\mu\text{A}/\text{lm}$  and high uniformity between each anode (Grégori et al. 2012). The data acquisition was performed by a custom-developed software package (Fig. 2) equipped with all the conventional flow cytometry analysis features (histograms, scatter plots, region gating, back-gating, basic statistics) as well as statistical processing for spectral data analysis (including principal component analysis and conversion of data vectors into hyperspherical coordinates).

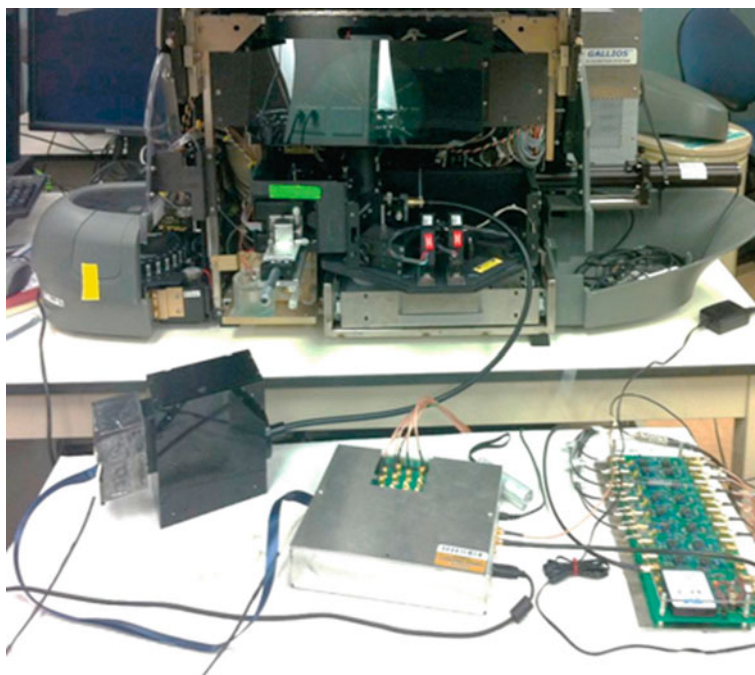
The second Purdue prototype, shown in Fig. 3, was developed using a modified FC500 flow cytometer (Beckman Coulter). The spectral detection module utilized a custom-enhanced version of Hamamatsu's high sensitive compact spectrometer (HSS A10766). The data acquisition system was upgraded to Beckman Coulter Gallios electronics capable of handling multichannel data collection. Experiments using fluorescent microspheres and lymphocytes labeled with a cocktail of antibodies (CD45/FITC, CD4/PE, CD8/ECD, CD3/Cy5) demonstrated the ability of the prototype to simultaneously collect 32 narrow bands of fluorescence from single particles flowing at about 1,000 events/second across the laser beam. The 32 discrete values collected at the single-cell level provide a proxy of the full fluorescence emission spectrum measured for each particle (Fig. 4).

Obviously, as the number of collected variables has increased, an advanced statistical processing is required in order to separate various clusters of cells analyzed in the spectral system (Fig. 5). The data analysis can be performed

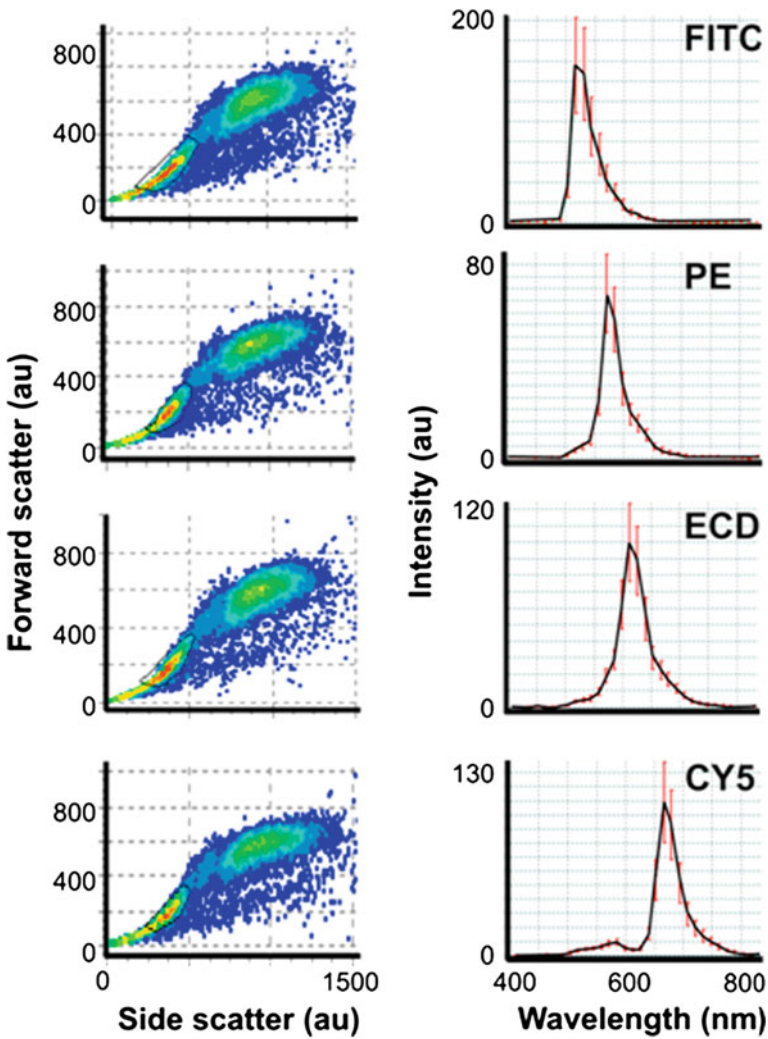




**Fig. 2** Screenshot of cytospec software package capable of acquisition and processing of hyperspectral cytometry data. The package was developed at PUCL by Valery Patsekina and is freely available at [http://www.cyto.purdue.edu/Purdue\\_software](http://www.cyto.purdue.edu/Purdue_software)

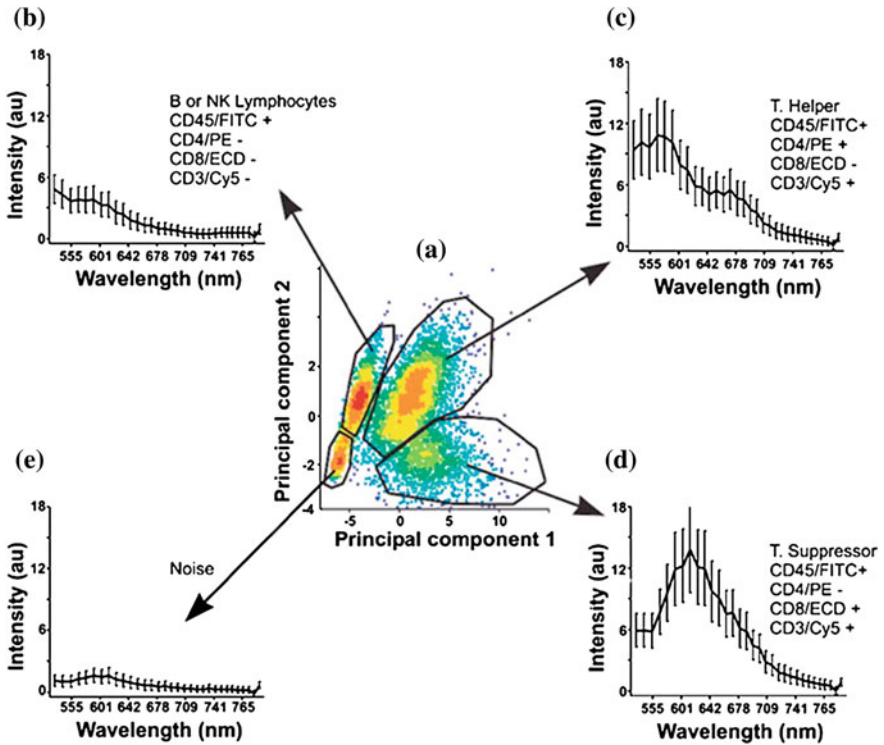


**Fig. 3** Photograph of a modified FC500 cytometer (Beckman Coulter) equipped with data acquisition subsystem from a Gallios instrument (Beckman Coulter) and the hyperspectral module developed at PUCL



**Fig. 4** *Left column*, forward versus side-scatter cytogram recorded for control samples consisting of blood labeled with a single antibody conjugated with FITC, PE, ECD, or Cy5. *Right column*, corresponding average spectrum (and standard deviation for each channel) obtained for each single-stained control from the spectra of  $\sim 7,000$  lymphocytes

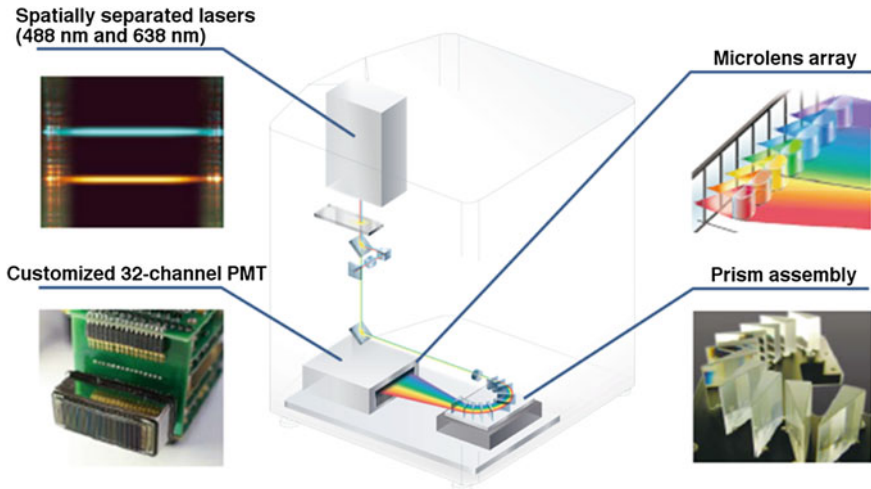
employing either linear unmixing techniques followed by gating or dimensionality reduction approaches paired with supervised classification (Grégori et al. 2012; Novo et al. 2013).



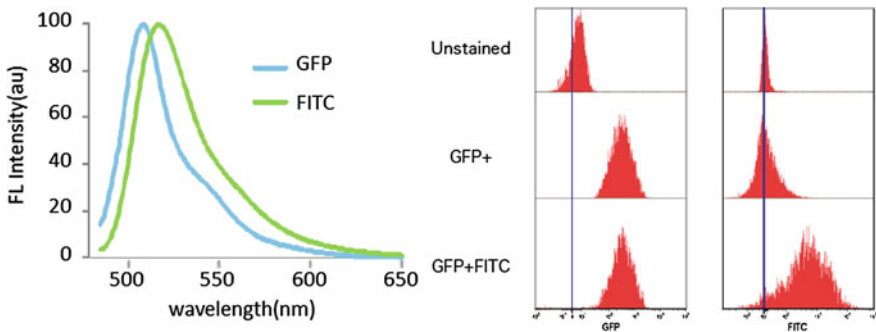
**Fig. 5** Different subpopulations of blood cells are clustered and classified on the basis of spectral data without the need for unmixing. Result of a principal component analysis performed on a flow cytometry data file corresponding to a blood sample incubated with a cocktail of antibodies labeled with four fluorochromes (FITC, PE, ECD, and Cy5). **a** Lymphocytes were first gated out from the rest of the particles on a forward-scatter versus side-scatter cytogram. In this example, four groups were discriminated based on the principal component analysis. The average spectrum (and the standard deviation per channel) is displayed for **b** B or natural killer lymphocytes, **c** T helper lymphocytes, **d** T suppressor lymphocytes, and **e** noise/detritus

### 5.2 Sony SP6800 Spectral Analyzer

The recently introduced Sony spectral analyzer offers a unique detection unit that captures all the emitted sample fluorescence ranging from 500 to 800 nm. This is achieved through a novel optical path utilizing a system of ten consecutive prisms and a custom-built 32-channel PMT, with independent voltage-controllable anodes (Fig. 6). Sample emission is directed through the prism set in order to disperse the signal into multiple separate bands (colors). A custom microlens array assembly then focuses each band of light onto a specific channel of the PMT array. This design limits the photon loss due to the presence of plates dividing the PMT channels and minimizes the crosstalk between the adjacent channels. The system is



**Fig. 6** Schematic of Sony SP6800 spectral cell analyzer. Spatially separated lasers (488 nm and 638 nm) excite the labeled cells flowing through the flow chamber chip. The fluorescence spectra of the sample in the range of 500–800 nm are collected by a 32-channel linear array PMT detector equipped with a multi-prism monochromator



**Fig. 7** Separation of GFP and FITC signals using Sony spectral system. *Data courtesy of W.C. Hyun (UCSF)*

capable of performing up to 15-color analyses with only two lasers (488 nm and 638 nm). With this functionality, the SP6800 has the ability to determine the spectral profile of autofluorescence from a cell and automatically remove it from a stained sample, improving signal-to-noise and data accuracy. Figure 7 shows an example of simultaneous detection and unmixing of spectra derived from adjacent fluorescent proteins and fluorochromes (such as GFP and FITC), which traditional flow cytometers cannot separate.

In addition, the SP6800 is capable of combining the channels in the PMT to form “virtual filters” or “virtual bands.” This sophisticated arrangement eliminates the need to use a secondary detection system of optical filter assemblies that are

employed in standard flow cytometers. This also allows the SP6800 to function as a regular polychromatic, multiparametric FC with the capability of fine-tuning the detection to adjust for particular staining strategies and for problems with label intensities or marker abundance. Another unique function, adapted from Sony's DVD and Blu-ray laser-tracking technology, allows the researcher to gather information about each sample component by determining its location within the flow cell. This, in turn, can be used in cytometry analysis to decrease signal variability and improve coefficients of variations for the whole sample or individual populations as are assessed during cell cycle analysis, for example.

## 6 Spectral Data Analysis

In addition to differences in hardware, a crucial difference between spectral cytometry and multiparametric polychromatic cytometry is in the approach to data processing, specifically signal unmixing. Traditional cytometry relies on a process known as compensation for correction of spectral overlap. As mentioned before, the wide collection bands of conventional cytometers and the broad emission spectra of organic fluorochromes lead to significant spectral overlap between the signals emitted by the various fluorochromes. The currently used compensation approach simply implements in a software format the hardware-based concept proposed by Parks in 1977 (Loken et al. 1977). In the simplest two-band setting the compensation circuitry is comprised of two differential amplifiers. Since one signal is routed through a potentiometer to the positive side of one amplifier and the negative side of the other, a fraction of one signal can be subtracted from the other signal and vice versa. The idea can be expressed mathematically as

$$\begin{cases} a_1 = r_1 - qa_2 \\ a_2 = r_2 - pa_1 \end{cases} \Rightarrow \begin{cases} r_1 = a_1 + qa_2 \\ r_2 = pa_1 + a_2 \end{cases}$$

where  $a_1$  and  $a_2$  are the abundances of fluorochromes 1 and 2, parameters  $q$  and  $p$  describe the proportion of the fluorochromes signal measured in "incorrect" detectors owing to spectral overlap, and  $r_1$  and  $r_2$  are the actual measurements of fluorochrome intensities.

In the vector/matrix convention the assumed process of signal formation and the subsequent compensation can be summarized as

$$\begin{bmatrix} r_1 & r_2 \end{bmatrix} = \begin{bmatrix} 1 & q \\ p & 1 \end{bmatrix} \begin{bmatrix} a_1 & a_2 \end{bmatrix} \Rightarrow \begin{bmatrix} a_1 & a_2 \end{bmatrix} = \begin{bmatrix} 1 & q \\ p & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 & r_2 \end{bmatrix}$$

The matrix  $\begin{bmatrix} 1 & q \\ p & 1 \end{bmatrix}$  describing the proportion of fluorescence intensity that is measured in a channel other than the channel dedicated to a particular fluorochrome is called the spillover matrix (denoted herein by **S**). The inversion of this

matrix is known as the compensation matrix  $\mathbf{C}$ . The “compensated” result can be easily found by multiplying the measured signal by the compensation matrix:

$$\begin{bmatrix} a_1 & a_2 \end{bmatrix} = \mathbf{S}^{-1} \begin{bmatrix} r_1 & r_2 \end{bmatrix} = \mathbf{C} \begin{bmatrix} r_1 & r_2 \end{bmatrix} = \begin{bmatrix} \frac{r_1 - qr_2}{1 - pq} & \frac{pr_1 - r_2}{1 - pq} \end{bmatrix}$$

Obviously, the compensation process can be generalized to any number of fluorochrome detectors (Bagwell and Adams 1993). In the matrix notation:

$$\begin{aligned} \mathbf{r} &\approx \mathbf{S}\mathbf{a} + \mathbf{n} \\ \mathbf{a} &\approx \mathbf{S}^{-1}\mathbf{r} = \mathbf{C}\mathbf{r} \end{aligned} \tag{1}$$

where  $\mathbf{r}$  denotes the vector of observations of length  $L$  (the number of detector channels/bands employed in the hyperspectral FC system),  $\mathbf{S}$  an  $L \times f$  spillover matrix ( $f$  being the number of labels used in an experiment),  $\mathbf{a}$  the vector of length  $f$  of fluorochrome abundances, and  $\mathbf{n}$  a vector of length  $L$  that denotes noise.

As illustrated above, the “compensated” signal (that is, a signal with a fraction of unwanted signals removed) can easily be found by inverting a spillover matrix, which is essentially a matrix representing spectra of all used fluorochromes normalized to the peaks (hence, one on the matrix diagonal). Owing to the spectral overlap, the broad spectra of the typical organic fluorochromes used as labels, and the increased number of spectral bands, the compensated signal becomes relatively smaller and smaller. The signal-to-noise ratio in the individual bands decreases as well.

The rationale behind compensation assumes two important arrangements in the experimental setting. Firstly, the number of fluorochromes is identical to the number of spectral channels (detectors) used in the experiment. A single band/detector is dedicated to measurement of signal from a single fluorochrome. In other words, the signal from a fluorochrome spilling over to other channels is treated as unwanted background and removed.

Hyperspectral cytometry differs substantially in the experimental arrangement, and consequently in the assumptions regarding signal formation, data collection, and signal unmixing. In hyperspectral cytometry data analysis, fluorescent labels are not considered to be necessarily linked with single dedicated detection bands. Every label emits a spectrum, which potentially can be detected in all the bands. The resultant signal measured from cells labeled with multiple fluorescent tags is a linear combination of these spectra:

$$\mathbf{r} = \mathbf{M}\mathbf{a} + \mathbf{n} \tag{2}$$

where  $\mathbf{M}$  is a spectral signature matrix (or mixing matrix), in which columns denote spectra of the fluorochromes used, and  $\mathbf{a}$  and  $\mathbf{n}$  are, as previously defined, related to fluorochrome abundance and noise, respectively.

This reasoning can be applied to flow cytometry data regardless of the number of spectral bands, and it follows a long—established methodology of spectral unmixing. The simplest version of the concept is well illustrated in the manuscript

by Bagwell and Adams (Bagwell and Adams 1993), as well as in many spectral imaging publications (Garini et al. 2006; Keshava and Mustard 2002; Nielsen 2001; Settle and Drake 1993; Zimmermann 2005). The spectral unmixing (dubbed “additive compensation” in the cytometry field) involves addition of all portions of unmixed signal originating from a given fluorochrome. Mathematically it uses a notation very similar to that of compensation:

$$\begin{cases} r_1 = pa_1 + (1 - q)a_2 \\ r_2 = (1 - p)a_1 + qa_2 \end{cases}$$

$$[r_1 \ r_2] = \begin{bmatrix} p & 1 - q \\ 1 - p & q \end{bmatrix} [a_1 \ a_2]$$

One may notice that the mixing matrix  $\mathbf{M}$  is columnwise normalized to 1. Consequently, even for a two-color arrangement, the spectral unmixing approach leads to a different result than does compensation:

$$\begin{aligned} [a_1 \ a_2] &= \begin{bmatrix} p & 1 - q \\ 1 - p & q \end{bmatrix}^{-1} [r_1 \ r_2] = \mathbf{M}^{-1} [r_1 \ r_2] \\ &= \begin{bmatrix} \frac{qr_1 + (q-1)r_2}{p+q-1} & \frac{pr_2 + (p-1)r_1}{p+q-1} \end{bmatrix} \end{aligned}$$

If the number of spectral bands is equal to the number of fluorochromes, the general solution can be expressed as

$$\hat{\mathbf{a}} = \mathbf{M}^{-1} \mathbf{r}$$

where  $\hat{\mathbf{a}}$  is an estimation of pure fluorochrome intensities, and hence their abundances. The number of spectral bands (channels) in a hyperspectral FC system is larger than the number of fluorochromes used. Therefore the mixing matrix  $\mathbf{M}$  is not square and it cannot be inverted.

Traditionally, the approach used in imaging applications utilizes a least-square technique (LS), which is equivalent to computation of Moore–Penrose pseudoinverse. This same method can be applied to FC data:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a} \in \mathbb{R}} \{ (\mathbf{r} - \mathbf{M}\mathbf{a})^T (\mathbf{r} - \mathbf{M}\mathbf{a}) \} \Rightarrow \hat{\mathbf{a}} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{r} \quad (3)$$

Unmixing the flow cytometry data using LS may result in negative abundances (i.e., unmixed signal lower than zero). As in imaging, this problem can be solved by imposing physical constraints on the unmixing process: a non-negativity constraint assures that all the results are positive, while the sum-to-one constraint states that all the unmixed values must sum to 100 % of the mixed input fluorescence signal.

$$\sum_{i=1}^p a_i = \mathbf{1}^T \mathbf{a} = \sum_{i=1}^p r_i = \mathbf{1}^T \mathbf{r}, \quad a_i \geq 0$$



The minimization of  $\|\mathbf{r} - \mathbf{M}\mathbf{a}\|_2$  with the additional constraints above must be performed using numeric methods, as there is no closed-form solution to constrained minimization.

The simple unmixing models, in the unconstrained version (Eq. 3) or with the additional constraints above, work relatively well in typical imaging applications, and they may be applied to spectral FC data without any further modifications. However, it must be pointed out that the described model implicitly assumed that noise in Eq. 2 is an additive, Gaussian noise. Consequently, the ordinary LS solution operates under an assumption of data homoskedasticity (homogeneity of variance). However, the noise model in fluorescence observation is actually Poisson-like. Therefore the variance of the fluorescence intensity measurements increases with the intensity, resulting in heteroskedastic data.

This information can be incorporated in the unmixing model, as recently demonstrated in a report by Novo, Grégori, and Rajwa (Novo et al. 2013). Starting from the assumption of shot noise—limited measurements, the authors showed that the heteroskedasticity can be accounted for in unmixing by utilizing Poisson regression process defined within generalized linear models:

$$\hat{\mathbf{a}} = \underset{\mathbf{a} \in \mathbb{R}_{\geq 0}}{\operatorname{arg\,min}} \{ 2\mathbf{j}^T (\mathbf{r} \circ \log(\mathbf{r} \circ \overline{\mathbf{M}\mathbf{a}}) - (\mathbf{r} - \mathbf{M}\mathbf{a})) \}$$

where  $\mathbf{j}$  is an  $L \times 1$  sum vector of 1,  $\overline{\mathbf{M}\mathbf{a}}$  is a Hadamard inversion of  $\mathbf{M}\mathbf{a}$ , and  $\circ$  denotes element-wise multiplication (Hadamard product).

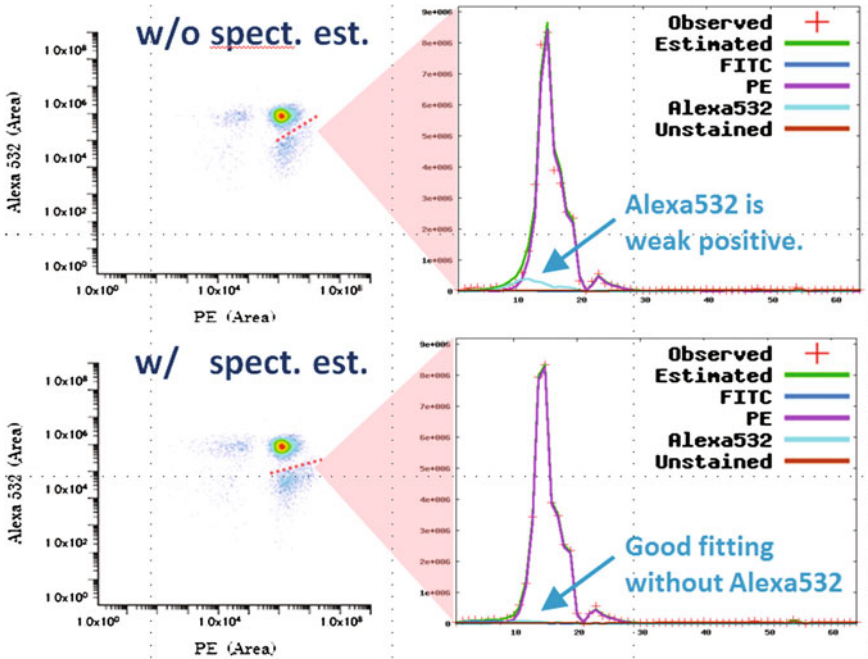
The methodologies described above assume that the mixing matrix  $\mathbf{M}$  is known and does not change. However, the spectrum of fluorochromes is not stable and may vary owing to multiple experimental factors. Therefore, in some experimental settings both  $\mathbf{a}$  and  $\mathbf{M}$  should be treated as unknowns, but this leads to the problem of decomposing the measured signals into two non-negative matrices. This issue, known as non-negative matrix factorization, has been widely studied in the context of blind source separation (Bilgin et al. 2012; Lee and Seung 1999; Pauca et al. 2006; Rabinovich et al. 2003).

In their data acquisition/processing software Sony implemented an algorithm described by Sekino et al. that estimates not only the fluorochrome abundances  $\mathbf{a}$  but also the mixing matrix  $\mathbf{M}$  based on probabilistic modeling and Bayesian estimation.<sup>3</sup> In this model, the spectrum of each fluorochrome is assumed to be generated from a prior distribution that corresponds to our belief regarding the likely shape of the spectrum. For example, one may use the truncated normal distribution to model the prior:

$$\mathbf{m}_i \sim N_{m_i \geq 0}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

<sup>3</sup> Masashi Sekino, Yasunobu Kato, Tatsumi Ito, “Probabilistic Spectrum Compensation for Flow Cytometry.” Presented at CYTO2012, XXVII Congress of International Society for Advancement of Cytometry, 23–27 June, 2012, Leipzig, Germany..





**Fig. 8** Dynamic estimation of the spectrum removes the artifact showing the presence of weak positive Alexa532 signal for cells labeled only with PE. After activation of the spectrum estimation technique the Alexa532 signal was correctly found to be approximately zero

Therefore,  $\mathbf{m}_i$  is normally distributed with mean vector  $\mu_i$  and the covariance matrix  $\Sigma_i$ . These parameters can be estimated from data collected on samples stained with a single label. Obviously  $\mathbf{m}_i$  should have only non-negative values. The measured signal is modeled by Eq. 2, with normally distributed error and the non-negativity constraint:

$$\mathbf{r} = \mathbf{M}\mathbf{a} + \mathbf{n}, \quad \mathbf{n} \sim N(0, \Lambda), \quad a_i \geq 0$$

The variational Bayesian algorithm enables an effective estimation of both  $\mathbf{M}$  and  $\mathbf{a}$  under these settings and constraints. Figure 8 demonstrates that the simultaneous estimation of  $\mathbf{M}$  improves chances of obtaining a relevant estimation of  $\mathbf{a}$ .

Finally, in some applications the multiplexing by spectral tags may not require spectral unmixing at all. In this setting it may be beneficial to classify the spectra directly, as opposed to classification based upon unmixed intensities. A large number of techniques may be utilized here, including unsupervised data reduction (using, for example, principal component analysis, independent component analysis, or factor analysis) or supervised techniques (such as neural networks or support vector machines) as illustrated in Fig. 5.

## 7 Potential Applications

The technology of spectral cytometry is still in its infancy; therefore, it would be premature to speculate on its potential impact on single-cell analysis or on the breadth and range of possible applications. However, it has already been demonstrated that the use of a spectral approach improves the ability to measure FRET-based molecular beacons used to identify specific stem-cell differentiation states (Bernstein and Hyun 2012). Generally, spectral analysis in cytometry is expected to improve the quality of all FRET-type measurements as has been shown in spectral microscopy applications (Zimmermann 2005; Zimmermann et al. 2002).

The number of bands collected in spectral cytometry experiments is larger than the number of fluorochromes used. This implies that spectral unmixing is always performed in an overdetermined setting. Consequently, accessing the spectral data one can unmix autofluorescence or any other background signal not originating from the labels of interest. This will lead to increases in measurement accuracy. As shown with spectral microscopy, acquisition of an entire spectral range allows for better separation of spectrally similar fluorescent proteins (Haraguchi et al. 2002; Zimmermann 2005; Zimmermann et al. 2003). Additionally, the presence of spectral information makes it possible to construct sophisticated unmixing strategies that take advantage of signal-formation models such as the recently proposed Poisson unmixing (Novo et al. 2013).

Finally, the use of spectral methodology can be advantageous if the spectra of the used labels change during the experiment. Mathematical techniques based on non-negative matrix factorization are capable of estimating both the label signals and the spectral matrices. Again, work developed for spectral imaging applications could be applied to spectral cytometry (Bilgin et al. 2012; Lee and Seung 1999; Pauca et al. 2006; Rabinovich et al. 2003).

## References

- Asbury CL, Esposito R, Farmer C, van den Engh G (1996) Fluorescence spectra of DNA dyes measured in a flow cytometer. *Cytometry* 24:234–242. doi:10.1002/(SICI)1097-0320(19960701)24:3<234:AID-CYTO6>3.0.CO;2-H
- Bagwell CB, Adams EG (1993) Fluorescence spectral overlap compensation for any number of flow cytometry parameters. *Ann N Y Acad Sci* 677:167–184. doi:10.1111/j.1749-6632.1993.tb38775.x
- Bernstein HS, Hyun WC (2012) Strategies for enrichment and selection of stem cell-derived tissue precursors. *Stem Cell Res Ther* 3:17. doi:10.1186/scri108
- Bilgin CC, Rittscher J, Filkins R, Can A (2012) Digitally adjusting chromogenic dye proportions in brightfield microscopy images. *J Microsc* 245:319–330. doi:10.1111/j.1365-2818.2011.03579.x

- Brown LO, Doorn SK (2008a) A controlled and reproducible pathway to dye-tagged, encapsulated silver nanoparticles as substrates for SERS multiplexing. *Langmuir ACS J Surf Colloids* 24:2277–2280. doi:[10.1021/la703853e](https://doi.org/10.1021/la703853e)
- Brown LO, Doorn SK (2008b) Optimization of the preparation of glass-coated, dye-tagged metal nanoparticles as SERS substrates. *Langmuir ACS J Surf Colloids* 24:2178–2185. doi:[10.1021/la703218f](https://doi.org/10.1021/la703218f)
- Buican TN (1990) Real-time fourier transform spectrometry for fluorescence imaging and flow cytometry. In: *Proceedings of SPIE 1205*. SPIE, Los Angeles, CA, pp 126–133
- Garini Y, Young IT, McNamara G (2006) Spectral imaging: principles and applications. *Cytometry A* 69A:735–747. doi:[10.1002/cyto.a.20311](https://doi.org/10.1002/cyto.a.20311)
- Gauci MR, Vesey G, Narai J et al (1996) Observation of single-cell fluorescence spectra in laser flow cytometry. *Cytometry* 25:388–393. doi:[10.1002/\(SICI\)1097-0320\(19961201\)25:4<388:AID-CYTO11>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-0320(19961201)25:4<388:AID-CYTO11>3.0.CO;2-R)
- Goddard G, Martin JC, Naivar M et al (2006) Single particle high resolution spectral analysis flow cytometry. *Cytometry A* 69A:842–851. doi:[10.1002/cyto.a.20320](https://doi.org/10.1002/cyto.a.20320)
- Grégori G, Patsekina V, Rajwa B et al (2012) Hyperspectral cytometry at the single-cell level using a 32-channel photodetector. *Cytometry A* 81A:35–44. doi:[10.1002/cyto.a.21120](https://doi.org/10.1002/cyto.a.21120)
- Haraguchi T, Shimi T, Koujin T et al (2002) Spectral imaging fluorescence microscopy. *Genes Cells Devoted Mol Cell Mech* 7:881–887
- Johnson PE, Lund ML, Shorthill RW et al (2001) Real time biodetection of individual pathogenic microorganisms in food and water. *Biomed Sci Instrum* 37:191–196
- Keshava N, Mustard JF (2002) Spectral unmixing. *SigProcess Mag IEEE* 19:44–57. doi:[10.1109/79.974727](https://doi.org/10.1109/79.974727)
- Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401:788–791. doi:[10.1038/44565](https://doi.org/10.1038/44565)
- Loken MR, Parks DR, Herzenberg LA (1977) Two-color immunofluorescence using a fluorescence-activated cell sorter. *J Histochem Cytochem* 25:899–907. doi:[10.1177/25.7.330738](https://doi.org/10.1177/25.7.330738)
- Nielsen AA (2001) Spectral mixture analysis: linear and semi-parametric full and iterated partial unmixing in multi- and hyperspectral image data. *J Math Imaging Vis* 15:17–37. doi:[10.1023/A:1011269530293](https://doi.org/10.1023/A:1011269530293)
- Nolan JP, Condello D, Duggan E et al (2013) Visible and near infrared fluorescence spectral flow cytometry. *Cytometry A* 83A:253–264. doi:[10.1002/cyto.a.22241](https://doi.org/10.1002/cyto.a.22241)
- Novo D, Grégori G, Rajwa B (2013) Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices. *Cytom Part J Int Soc Anal Cytol* 83:508–520. doi:[10.1002/cyto.a.22272](https://doi.org/10.1002/cyto.a.22272)
- Pauca VP, Piper J, Plemmons RJ (2006) Nonnegative matrix factorization for spectral data analysis. *Linear Algebra Its Appl* 416:29–47. doi:[16/j.laa.2005.06.025](https://doi.org/10.1016/j.laa.2005.06.025)
- Perfetto SP, Chattopadhyay PK, Roederer M (2004) Seventeen-colour flow cytometry: unravelling the immune system. *Nat Rev Immunol* 4:648–655. doi:[10.1038/nri1416](https://doi.org/10.1038/nri1416)
- Rabinovich A, Agarwal S, Laris CA, et al (2003) Unsupervised color decomposition of histologically stained tissue samples. *Adv Neural Inf Process Syst* 16:667–674
- Robinson JP (2004) Multispectral cytometry: the next generation. *Biophotonics Int* 11:36–40
- Robinson JP, Rajwa B, Gregori G et al (2005) Multispectral cytometry of single bio-particles using a 32-channel detector. *Proc SPIE* 5692:359–365
- Robinson JP, Rajwa B, Grégori G, et al (2007) Multi-spectral detector and analysis system. US Patent 7280204 B2
- Roederer M (2001) Spectral compensation for flow cytometry: Visualization artifacts, limitations, and caveats. *Cytometry* 45:194–205
- Roederer M, De Rosa S, Gerstein R et al (1997) 8 color, 10-parameter flow cytometry to elucidate complex leukocyte heterogeneity. *Cytometry* 29:328–339
- De Rosa SC, Herzenberg LA, Herzenberg LA, Roederer M (2001) 11-color, 13-parameter flow cytometry: identification of human naive T cells by phenotype, function, and T-cell receptor diversity. *Nat Med* 7:245–248. doi:[10.1038/84701](https://doi.org/10.1038/84701)

- Settle JJ, Drake NA (1993) Linear mixing and the estimation of ground cover proportions. *Int J Remote Sens* 14:1159. doi:[10.1080/01431169308904402](https://doi.org/10.1080/01431169308904402)
- Shapiro HM (2003) *Practical Flow Cytometry*, 4th edn. Wiley-Liss, Hoboken, NJ, USA
- Steen HB (1986) Simultaneous separate detection of low angle and large angle light scattering in an arc lamp-based flow cytometer. *Cytometry* 7:445–449. doi:[10.1002/cyto.990070509](https://doi.org/10.1002/cyto.990070509)
- Stokke T, Steen HB (1986) Fluorescence spectra of Hoechst 33258 bound to chromatin. *Biochim Biophys Acta BBA - Gene Struct Expr* 868:17–23. doi:[10.1016/0167-4781\(86\)90081-3](https://doi.org/10.1016/0167-4781(86)90081-3)
- Wade CG, Rhyne RH Jr, Woodruff WH et al (1979) Spectra of cells in flow cytometry using a vidicon detector. *J Histochem Cytochem Off J Histochem Soc* 27:1049–1052
- Wang J-CE, Kobie JJ, Zhang L et al (2009) An 11-color flow cytometric assay for identifying, phenotyping, and assessing endocytic ability of peripheral blood dendritic cell subsets in a single platform. *J Immunol Methods* 341:106–116. doi:[10.1016/j.jim.2008.11.002](https://doi.org/10.1016/j.jim.2008.11.002)
- Watson DA, Brown LO, Gaskill DF et al (2008) A flow cytometer for the measurement of Raman spectra. *Cytometry A* 73:119–128. doi:[10.1002/cyto.a.20520](https://doi.org/10.1002/cyto.a.20520)
- Zimmermann T (2005) *Spectral imaging and linear unmixing in light microscopy*. Microsc. Tech. Springer Verlag, Berlin/Heidelberg, pp 245–265
- Zimmermann T, Rietdorf J, Girod A et al (2002) Spectral imaging and linear un-mixing enables improved FRET efficiency with a novel GFP2-YFP FRET pair. *FEBS Lett* 531:245–249
- Zimmermann T, Rietdorf J, Pepperkok R (2003) Spectral imaging and its applications in live cell microscopy. *FEBS Lett* 546:87–92. doi:[10.1016/S0014-5793\(03\)00521-0](https://doi.org/10.1016/S0014-5793(03)00521-0)