Cytometry

Quadratic Form: A Robust Metric for Quantitative Comparison of Flow Cytometric Histograms

Tytus Bernas,[1,2] Elikplimi K. Asem,[3] J. Paul Robinson,[2] Bartek Rajwa[2]*

[1]Department of Plant Anatomy and Cytology, Faculty of Biology and Protection of Environment, Jagiellonska 28, Katowice, Poland

[2]Purdue University Cytometry Laboratories, 1203 West State St., West Lafayette, Indiana

[3]Purdue University School of Veterinary Medicine, 625 Harrison St., West Lafayette, Indiana

This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/1552-4922/suppmat.

*Correspondence to: Bartek Rajwa, Purdue University Cytometry Laboratories, 1203 West State St., West Lafayette, IN, USA.

Email: brajwa@purdue.edu

• **Abstract**

Comparison of fluorescence distributions is a fundamental part of the analysis of flow cytometric data. This approach is applied to detect differences between control and test sample and thus analyze a biological response. Comparison of standard test samples over time provides an estimate of instrument stability for quality control. However, application of statistical methods of distribution comparison in flow cytometry is difficult owing to instrument noise and the complex shape of intensity distributions. We applied quadratic form (QF) as a mathematical metric for comparison of flow cytometry histograms. QF operates on histograms as vectors and calculates the total distance in an interbin manner using a matrix of distances between single histogram bins. Euclidean interbin distance and histograms normalized to unity were used. Critical values corresponding to 95% significance level were calculated using Monte-Carlo simulation and single-maximum Gaussian distributions populated with several numbers of events. The QF statistic was then validated for non-Gaussian single-maximum distributions and multiple-maxima distributions. We determined that the critical values for Gaussian distributions depended on standard deviations and number of events in the compared histograms. A simple empirical function was constructed to characterize this dependence. Furthermore, it was verified that critical values (corresponding to 95% significance) for non-Gaussian histograms were similar to values for the Gaussian histograms characterized by the same standard deviation. We applied the QF statistic to estimate the differences between histograms of DNA content (ploidy) in cells of old and young leaf tissue of *Brassica campestris.* Furthermore, we quantified differences in fluorescence intensity in immunostaining of human lymphocytes. Quadratic form (QF) provides a true (mathematical) metric for estimation of distance between flow cytometry histograms of arbitrary shape. QF can be applied as a statistical test for estimation of significance of the distance measure. The respective critical values depend only on the number of events and standard deviations of compared histograms and are not affected by distribution shape. Therefore, applications of QF do not require assumptions concerning distribution shape and can be easily implemented in practice. This notion was confirmed using empirical distributions of DNA content in plant tissue and distributions of immunofluorescence in human cells. © 2008 International Society for Advancement of Cytometry

• **Key terms**

flow cytometry; data analysis; quadratic form; distance metric; histogram comparison; statistics

COMPARISON of fluorescence intensity distributions is a common task in the visualization of flow cytometry data and the most basic and rudimentary form of analysis of such data. Histograms are compared to verify reproducibility of sample measurements and to measure differences in cellular populations. Usually only the simplest descriptors of population, such as median, mean, geometric mean, and variance, are employed to characterize flow cytometry histograms. However, an accurate comparison of any pair of histograms should involve an application of a reproducible measure of (dis)similarity, and an estimation of the statistical significance of such a measure.

Table 1. Desirable properties of distance functions

	CONDITION		DESCRIPION	
1.	$D(x,y) \geq 0$		Non-negativity	⎫ Positive definiteness
2.	$D(x,y) = 0 \Leftrightarrow x = y$		Identity of indiscernibles	⎭
3.	$D(x,y) = D(y,x)$		Symmetry	
4.	$D(x,z) \leq D(x,y) + D(y,z)$		Subadditivity	
5.	$D(x,y) = D(x+a,y+a)$		Translational invariance	

Cytometrists employ two different approaches to achieve these goals. The first involves constructing parametric models of the population histograms and using an estimate of % positive (or negative) cells (for specific markers) or calculating a distance in parameter space (usually Euclidean). However, building such a cell population model requires detailed information (not always available) about the biology of the systems being studied. Furthermore, owing to the large number of parameters, estimation of histogram similarity using computational models is not always straightforward. The second approach employs some form of distance function based on nonparametric tests such as Kolmogorov–Smirnov (KS) (1–6), χ^2 (7–9), or Cramer–von Mises statistic (9,10). However, these methods are known to be overly sensitive to the shape of cytometry histograms and do not perform well when levels of freedom (number of histogram bins) are numerous and populated with small numbers of events. Consequently, these tests tend to underestimate the probability of uniqueness of discrete data sets.

To address these problems, an alternative method based on probability binning (PB) was introduced by Roederer

(8,11) and subsequently revised by Baggerly (12). PB-χ^2 uses χ^2 statistic with a modified binning such that it minimizes the maximal expected variance. PB-χ^2 requires that an intensity histogram of a studied sample be compared with an appropriate control. The control and the sample must be registered under the same measurement conditions since characteristics of the control histogram are used to construct a binning function for the sample histogram. In other words, the binning function (and consequently the distance function) is specific for every control/sample pair. Therefore, the PB-χ^2 distance measure is not quantitative in the mathematical sense, since it is not subadditive and does not provide ground distance estimation (see Tables 1 and 2).

Naturally, histogram comparison tasks are not limited to flow cytometry. Numerous histogram similarity measures have been applied in the image-processing domain (13–18). These include measures operating on single bins ("bin-to-bin" methods, such as the aforementioned KS or χ^2, but also Kullback–Leibler divergence or Bhattacharyya distance) and multiple bins ("cross-bin" distances). The latter group is comprised of quadratic form with dissimilarity-matrix (QF) and

Table 2. Summary of various dissimilarity measures discussed in this report

DISTANCE FUNCTION	SYMMETRY	SUBADDITIVITY	EMPLOYS GROUND DISTANCE FUNCTION	MULTIVARIATE HISTOGRAMS	COMP. COMPLEXITY	FORMULA
K-S	Yes	Yes	Yes	No	Low	$D(\mathbf{h}, \mathbf{f}) = \max\limits_{i=1}^{n} \lvert F(h)_i - F(f)_i \rvert$ where F($\mathbf{h}$) is a cumulative histogram of $\mathbf{h}$.
χ^2	Yes	No	No	Yes	Low	$D(\mathbf{h}, \mathbf{f}) = \sum\limits_{i=0}^{n} \frac{(h_i - f_i)^2}{(h_i + f_i)}$
QF	Yes	Yes	Yes	Yes	Medium	$D(\mathbf{h}, \mathbf{f}) = \sqrt{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} a_{ij}(h_i - f_i)(h_j - f_j)},$ where $a_{ij} = \sqrt{1 + (i - j)^2}$
EMD	Yes	Yes	Yes	Yes	High	$D(\mathbf{h}, \mathbf{f}) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} g(h_i, f_j) d(h_i, f_j)}{\sum_{i=1}^{n} \sum_{j=1}^{n} g(h_i, f_j)},$ where $d(h_i, f_j)$ denotes dissimilarity between bins i and j of histograms $\mathbf{h}$ and $\mathbf{f}$, and $g(h_i, f_j)$ is the cost of optimal mass transportation between the histograms.

L_n-Wasserstein distance (earth mover's distance, EMD) (17,18). These two nonparametric measures are true metric functions (i.e., they satisfy the conditions of positive definiteness, symmetry, and subadditivity), and are able to provide a distance estimation between multidimensional histograms. The EMD is the most advanced, but also the most unusual metric since it defines the distance computation between distributions as a solution to the Monge–Kantorovich mass-transportation problem. However, the complexity of EMD is larger than $O(N^3)$, where N is the number of histogram bins. Table 2 summarizes the distance functions used to compare histograms in flow cytometry or image processing.

In this report, we apply QF with (dis)similarity matrix as a metric for comparison of flow cytometry histograms. We also demonstrate how to calculate appropriate critical values (i.e., the values that our similarity test must exceed in order for the null hypothesis to be rejected) to establish confidence levels associated with such a comparison. Hence, we establish QF not only as a cytometry histogram (dis)similarity metric (distance function), but also as a statistic. Subsequently, we validate the QF statistic for Gaussian and several non-Gaussian distributions. To do so, we create a set of synthetic histograms, i.e., histograms of predefined shapes, based on randomly drawn subpopulations of events generated in silico. By comparing various histograms based on subpopulations drawn from a known population, we are able to show the sensitivity and relevance of the proposed measures of histogram similarity. Finally, we use fluorescence intensity distributions of real biological populations to demonstrate the applicability of the defined distance function and statistic to cases involving non-Gaussian histograms. Consequently, we demonstrate that QF distance can be reproducibly measured and expressed in a range of magnitudes (i.e. on an interval scale), providing a tool for quantitative comparison of cytometry histograms.

MATERIALS AND METHODS

Computer Hardware and Software

All computations and Monte–Carlo simulations were executed using Matlab R13 (Math Works) running under MS Windows 2000 (SP4) on a AMD Athlon XP 2800 + (1,950 MHz) machine equipped with 1 GB DDR RAM (333 MHz).

Flow Cytometry

Nuclei isolated from young and old leaf tissue of *Brassica campestris* (Goldball) were stained using DAPI (1 μM). Fluoresence of nuclear DNA stained with DAPI (ex. 405 nm, em. 450–480 nm) was measured using a Partec PAS III flow cytometer. Whole human blood lymphocytes were stained with Beckman–Coulter CD14-PE/CD45-FITC antibodies (dilutions from 1× to 512×). Fluorescence of FITC (ex. 488 nm, em. 505–545 nm) was measured with a Beckman–Coulter Cytomics FC 500 flow cytometer. The intensity was registered with 10-bit precision (1,024 channels) on a linear scale (DAPI) or logarithmic scale (FITC).

Construction of Histogram Metric

A quadratic form was used as used as a metric of the distance between two histograms (11). The QF distance D was calculated using the following formula:

$$D^2(\mathbf{h}, \mathbf{f}) = (\mathbf{h} - \mathbf{f})^T \mathbf{A}_j^i (\mathbf{h} - \mathbf{f}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(h_i - f_i)(h_j - f_j),$$

$$(1)$$

where $\mathbf{h}$ and $\mathbf{f}$ are vectors that list counts corresponding to each of the histogram bins (channels). These vectors can be normalized so that $0 \leq \mathbf{f}, \mathbf{h} \leq 1$ and $\sum_i h_i = \sum_i f_i = 1$. The i and j are histogram bin (channel) numbers in $\mathbf{h}$ and $\mathbf{f}$, respectively, and $\mathbf{A}_j^i = [a_{ij}]$ is the matrix of distances between the ith and jth bins. The distance (or dissimilarity) matrix was defined as $\mathbf{A}_j^i = \sqrt{1 + (i - j)^2}$. Alternatively, $\mathbf{A}_j^i$ could be also defined as a similarity matrix $1 - \sqrt{(i - j)^2 / d_{\max}}$, where $d_{\max}$ is a maximum distance (dissimilarity) between bins. A number of other (dis)similarity distances can potentially be used for QF calculation. In fact, one may calibrate a specific distance for a given biological experiment to ensure linear correspondence between biological phenomena described by a given change in histogram and calculated distance. A simple example of such a flexible ground distance matrix is $\mathbf{A}_j^i$ with $a_{ij} = -\frac{c}{1-c} + \frac{1}{1-c}\exp\left(-\beta\left(\frac{i-j}{d_{\max}}\right)^2\right)$ where β is a positive constant, and $c = \exp(-\beta)$ is a normalization factor. The histograms $\mathbf{h}$ and $\mathbf{f}$ had 1,024 bins in all the calculations. One may note that the Euclidean distance is a specific case of general QF distance where $\mathbf{A}_j^i = \mathbf{1}$ (the identity matrix). Interestingly, χ^2 (and PB-χ^2) distance is also a special case of QF for a specific $\mathbf{A}_j^i$ with $a_{ij} = \frac{1}{\sqrt{h_i h_j}}$ if $i = j$, and $a_{ij} = 0$, otherwise. Examples of various $\mathbf{A}_j^i$ matrices can be seen in Figure 1.

As dissimilarity metric for histograms (distributions) QF is a smooth and monotonic function of 3rd and 4th moments (skewness and kurtosis, respectively), which characterize distribution shape. This notion is illustrated by Supplemental Figure 1 (available online), which shows QF distances to a control histogram characterized by zero skewness and kurtosis excess and but identical mean and variance (i.e., Gaussian histogram). To ensure mutual independence of the moment changes, density templates were constructed using the Hermitte series method proposed in (19) and (20). This method involves reconstruction of a density template (probability density function, PDF) from the series of distribution moments.

Confidence Levels and Validation of QF Statistic

By definition, a distance function must satisfy the principle of identity of indiscernibles: $D(\mathbf{h}, \mathbf{f}) = 0 \Leftrightarrow \mathbf{h} = \mathbf{f}$. However, owing to natural biological variability, the distance between two histograms $\mathbf{h}$ and $\mathbf{f}$ drawn randomly from the same biological population is never zero. Therefore, a distance below which two histograms are considered to be similar

Figure 1. Examples of 10 × 10 (dis)similarity matrices $\mathbf{A}_j^i$ used in three different QF distance metrics, which can be used to compare 10-bin histograms: (**A**) $a_{ij} = \exp\left(-\beta\left(\frac{i-j}{d_{\max}}\right)^2\right)$ $\beta = 1$; (**B**) $a_{ij} = 1 - \sqrt{(i-j)^2/d_{\max}}$ where $d_{\max}$ is a maximum distance (dissimilarity) between bins, and (**C**) $a_{ij} = 1$ for $i = j$ and 0, otherwise (Euclidean distance).

enough to be treated as most likely drawn from one population has to be estimated. A Monte–Carlo algorithm was employed to find the 95% confidence levels for similarity of two such histograms. First, 17 discrete Gaussian intensity distributions Y_g (templates, normalized to unity) were generated:

$$Y_g(\mu, \sigma, l) = \frac{1}{\sigma\sqrt{2\pi}} e^{(1-\mu)^2/2\sigma}, \qquad (2)$$

where σ is the standard deviation, μ is the distribution average, and l represents the number of a histogram bin (0–1,023). Each of these synthetic distributions of different σ but centered in the middle of the scale was used as a probability density function to generate empirical histograms populated with several numbers of events (from 10,000 to 90,000). Two hundred fifty such histograms have been created for each of the templates and each tested number of events. This process represents a computer simulation of an actual cytometry measurement, for which a histogram of events represents a biological population. The pairs of histograms derived from one template distribution (a simulation of two separate cytometric measurements of one sample) were compared using the QF metric defined in Eq. (1). The resulting distances formed another distribution. 250 × [250 − 1]=62,250 comparisons were performed for each pair of defined event numbers and each template. Significance levels S_p^D for the constructed metric were defined as functions of number of events (N_1, N_2) and distribution width (σ):

$$S_p^D(N_1, N_2, \sigma) = \text{Perc}_p(D) \qquad (3)$$

where σ is the standard deviation associated with $\mathbf{h}$ and $\mathbf{f}$ [Eq. (1)], p is the significance level (0.95 or 0.99), $\mathbf{D}$ is the vector of all D values (distribution of distances), N_1 and N_2 are numbers of events in $\mathbf{h}$ and $\mathbf{f}$, and Perc_p is the pth percentile (e.g., 95th) of the distribution of distances.

The S_p^D significance levels were plotted against the respective standard deviation (σ) values for each pair of event numbers. An empirical dependence was constructed using nonlinear regression:

$$S_p^D(N_1, N_2, \sigma) = U\sigma^W, \qquad (4)$$

where U and W are fit coefficients.

To validate performance of the QF statistic for non-Gaussian distributions with a single maximum (see Fig. 2), a generalized log-normal formula was used:

$$Y_{ng}(v, \delta, l) = \frac{1}{\delta\sqrt{2\pi}} e^{[p\times\log(l)-v]^2/2\delta}, \qquad (5)$$

where δ is the standard deviation, v is the distribution average, l is the number of a histogram bin (0–1,023), and p is a power coefficient (1 for standard and for 3 modified log-normal distributions, respectively).

The v values were chosen so the positions of the maxima of these non-Gaussian distributions matched the center (μ) of
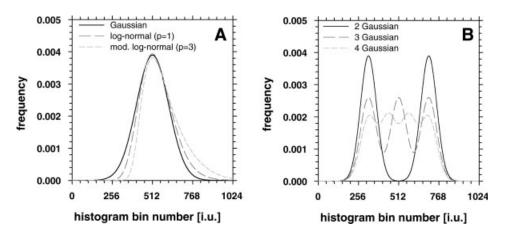
Figure 2. (**A**) Examples of single-maximum log-normal (long dash line) and modified log-normal (short dash line) template distributions. A Gaussian distribution (solid line) is shown for comparison. All the distributions have a maximum of 512 and a standard deviation of 102.5. (**B**) Examples of two-maxima (solid line), three-maxima (long dash line), and four-maxima (short dash line) Gaussian template distributions. The component (Gaussian) distributions (see Materials and Methods) had a standard deviation of 51.2 and separation of 192 (see Materials and Methods).

their Gaussian counterparts. One should note that in contrast to the latter, the former distributions were characterized by positive skewness (0.25–0.53 for standard and 0.56–0.66 for modified log-normal, respectively). Similarly, these distributions exhibited non-zero kurtosis excess (0.16–0.47 for standard and 0.27–0.43 for modified log-normal, respectively). Furthermore, performance of the statistic for distributions with multiple maxima was tested (see Fig. 2B). These distributions were generated using a sum of 2, 3, and 4 Gaussians as templates (normalized to unity). The means of component Gaussians [μ, calculated using Eq. (2)] were separated from one another by 64, 128, or 192 bins to mimic various degrees of overlap of subpopulations in flow cytometry histograms. Since these distributions were symmetric the skewness was 0. On the other hand values of kurtosis excess varied between -1.19 and -1.85 (depending on shape). As before, these templates were randomly populated with 10,000 to 90,000 events. The significance levels were calculated using Eq. (3)) and plotted against the respective distribution widths.

Application of QF to Study Changes in *Brassica* Leaf Ploidy

Nuclei were isolated from samples of young and old leaf tissue in *Brassica campestris* (Goldball) using a procedure described previously (21). DNA in isolated nuclei was stained with DAPI. Fluorescence intensity was measured using a flow cytometer. The experiments were repeated five times. From 8,000 to 10,000 nuclei were analyzed in each experiment. The flow histograms were averaged, filtered using a median filter (size 3), and normalized to unity to create reference fluorescence intensity distributions representing young and old leaf tissue, respectively. Synthetic template distributions for tissue of intermediate age were created by weighted averaging of young and old leaf reference distributions. The histograms representing randomly drawn subpopulations of the measured

samples were built by populating the template distributions with 10,000–90,000 events. These histograms were compared in pairs using the QF metric. In addition, the histograms corresponding to each of two reference distributions (i.e., corresponding to either young or old leaf tissue) were compared using PB-χ^2 (8), modified PB-χ^2 (12) (with 10 or 25 bins), and KS (1,2,22) statistics.
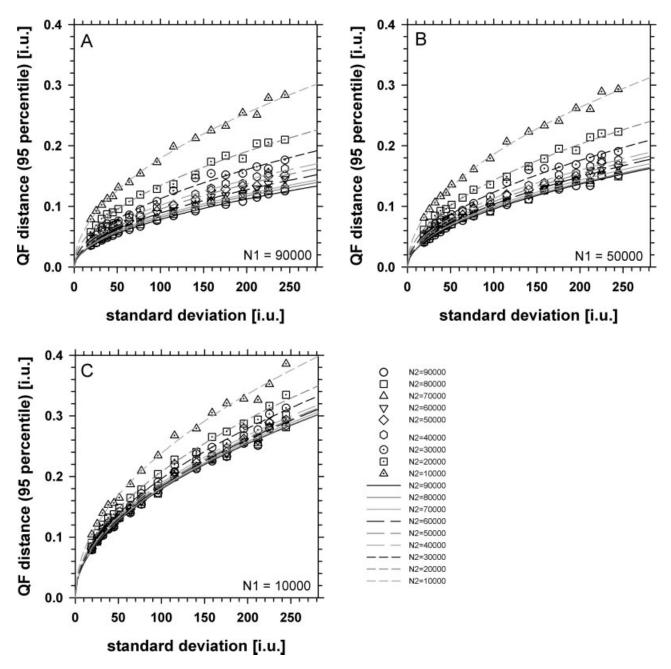
Application of QF for CD45 Lymphocyte Immunoflourescence Quantification

Whole blood was obtained from a healthy donor and stained with CD45-FITC antibodies (dilutions from 1× to 512×). Fluorescence of FITC was measured using a flow cytometer. 10,000 cells were registered for each experiment. The flow histograms were filtered using a median filter (size 3) and normalized to unity to create template fluorescence intensity distributions representing populations stained with different concentrations of antibody. The histograms representing randomly drawn subpopulations of the measured samples were built by populating the template distributions with 10,000–90,000 events. Histograms representing lymphocytes stained with antibody of intermediate dilutions (from 2× to 256×) were compared in pairs using QF metric to histograms representing strongly (1×) and weakly (512×) fluorescent cells. In addition, the histograms corresponding to either of these latter populations were compared using PB-χ^2 (8), modified PB-χ^2 (12) (with 10 or 25 bins), and KS (1,2,22) statistics.

Results

Calculation of QF Critical Values for Gaussian Distributions

Synthetic Gaussian-shaped histograms populated with numbers of events ranging from 10,000 to 90,000 were com-

Figure 3. Dependence of critical values (expressed as 95th percentile) of QF distance on standard deviation of Gaussian histograms (see Materials and Methods). The first group of compared histograms (h_1) were populated with 90,000 (**A**), 50,000 (**B**), and 10,000 (**C**) events, the second (h_2) with 10,000–90,000 events (indicated with respective symbols). Fit curves are shown using solid lines (70,000–90,000 h_2 events), long dash lines (40,000–60,000 h_2 events), and short dash lines (10,000–30,000 h_2 events).

pared in pairs using QF metric. Critical values were established as the 95th percentile of the distribution of QF distances obtained from comparison of pairs of histogram built from one model distribution (see Materials and Methods). In other words, the 95th percentile of QF distances measured between histograms representing a single known population (template distribution) is considered to be the value that the proposed statistic must exceed in order for the null hypothesis (that the histograms are similar) to be rejected. Not surprisingly, the

critical values increased with the standard deviation of the template distribution (Fig. 3). The values also increased with decreasing number of events comprising compared histograms (compare Figs. 3A, 3B, and 3C and different symbols in each figure). The increase in the critical values with standard deviation could be described using a power function of standard deviation (lines in Fig. 3 and Supplemental Fig. 2). The multiplicative-fit coefficient [U, Eq. (4)] increased with decreasing number of events in the two compared histograms (Supple-

mental Fig. 2A). The power-fit coefficient W [See Eq. (4)] remained constant in the function of number of events and standard deviation (Supplemental Fig. 2B). One should note that fit errors were small and therefore the critical values could be calculated in a simple and precise manner for any number of events.

Validation of Critical Values of QF Statistics for Non-Gaussian Distributions with a Single Maximum

To validate the QF statistic for non-Gaussian distributions with a single maximum, a series of synthetic standard log-normal and modified log-normal [see Eq. (5)] distributions (Fig. 2A) was generated. The synthetic histograms containing 10,000–90,000 events built were from these distributions, and the critical values of the QF statistics were calculated as described earlier. As with the Gaussian case, the values increased with decreasing number of events, and with increasing standard deviation (compare symbols in Figs. 4A, 4C, and 4E). Similar dependence was obtained when the critical values of QF statistics were calculated for modified log-normal distributions (Figs. 4B, 4D, and 4F). One should note that these dependences for non-Gaussian histograms could be accurately described using functions originally calculated for their Gaussian counterparts (populated with the same numbers of events), as shown by the lines in Figure 5. The only exceptions were slightly lower values obtained for modified log-normal histograms when simultaneously the width of the template distribution was large and the number of events small (Fig. 4F). Nonetheless, one may postulate that shapes of single-maximum histograms do not influence critical values of the proposed QF statistic.

Validation of Critical Values of QF Statistics for Distributions with Multiple Maxima

To validate the QF statistic for distributions with multiple maxima, a series of populations constructed as combinations of 2, 3, and 4 Gaussian distributions (Fig. 2B) was generated. As before, for the purpose of QF critical-values calculations, histograms containing 10,000–90,000 events were built from these populations. Once again, the critical values increased with decreasing number of events and with increasing standard deviation (compare symbols in Fig. 5 in series ADG, BEH, and CFI). One may note that the standard deviation was affected more by the separation distance of the component Gaussians than by their width. The critical values for these multiple-maxima histograms could be accurately described using functions calculated for their Gaussian counterparts (populated with the same numbers of events), as shown by the lines in Figure 5. Therefore, one may postulate that the number of histogram maxima does not influence critical values of the proposed QF statistic. Consequently, it may be concluded that the QF statistic is robust with respect to distribution shape and may be used to compare arbitrary histograms.
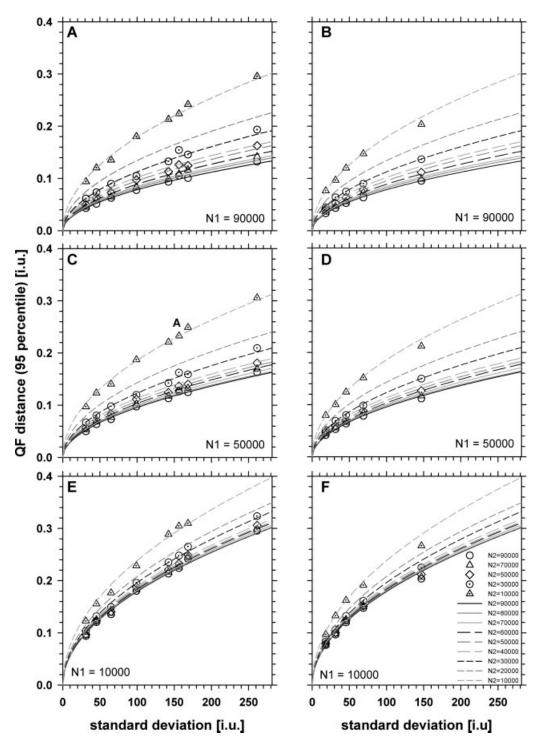
Application of QF to Measure Ploidy Changes in Development of Leaf Tissue

Practical application of the QF metric and statistic was demonstrated using DNA content histograms of nuclei from leaf tissue of *Brassica campestris* (Supplemental Fig. 3). As expected, the QF distance between histograms (generated in silico, but with biological templates) representing old tissue control and mixed tissue population increased with the contribution of the control in the mixed distribution (compare light and dark symbols in Fig. 6). The opposite effect (a decrease in QF distance) was observed when young tissue measurement was taken as the reference, and compared with the mixed distribution. One should note that the symbols indicating positions of distributions of mixed populations form straight lines in the space where the basis is formed by reference distributions of old tissue (x axis) and young tissue (y axis), respectively. Furthermore, positions of the symbols were similar when intermediate distributions were populated with large (Fig. 6A), moderate (Fig. 6B), and low (Fig. 6C) numbers of events. The QF critical values increased with decreasing number of events (compare panels A, B, and C in Fig. 6). Consequently, the minimal significant difference (in the sense of QF metric) between histograms increased with decreasing number of events.

In contrast, when the PB-χ^2 statistic was used to perform the task of comparison, it detected a difference (at 95% significance level, data not shown) in 0.001% to 8.4% cases of two histograms representing randomly chosen measurements of biological particles from one and the same population. This fraction of different histograms was affected by the histogram shape, numbers of events comprised in the histograms, and number of PB bins. One should note that at the 95% significance level the expected fraction is 5%. Therefore, depending on the data and the procedure of hypothesis testing, PB-χ^2 statistics underestimated or overestimated probability of samples drawn from the same population being different (i.e., exhibited false negative or false positive bias). Modified PB-χ^2 eliminated false positive bias and reduced the bias range of PB. However, this statistic exhibited false negative bias (fraction of 1.2%–3.1% of two histograms representing the same population recognized as different at 95% significance level). This effect also occurred also when KS was used instead of modified PB-χ^2 (the respective fraction was 0.06%–2.1%).
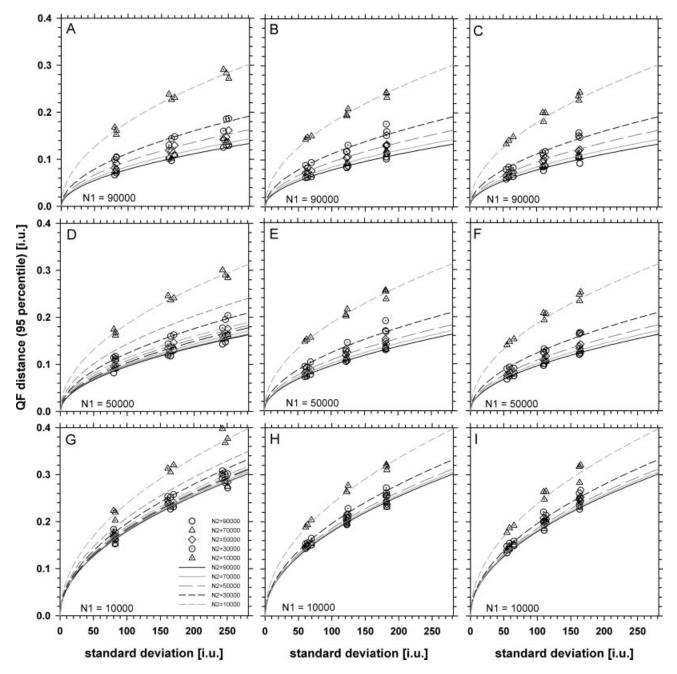
Application of QF for CD45 Lymphocyte Immunoflourescence Quantification

Practical application of the QF metric and statistic were also demonstrated using immunofluorescence from human lymphocytes immunostained with CD45-FITC antibody (Supplemental Fig. 7). As expected, the QF distance between histograms representing staining with concentrated antibody and its subsequent dilutions increased with decreasing concentration (compare light and dark symbols in Fig. 7). The opposite effect (a decrease in QF distance) was observed when measurement of the most weakly stained population (antibody dilution 512×) was taken as the reference. One

Figure 4. Dependence of critical values (expressed as 95th percentile) of QF distance on standard deviation of log-normal (**A,C,E**) and modified log-normal (**B,D,F**) histograms (see Materials and Methods). The first group of the compared histograms (h_1) were populated with 90,000 (A,B), 50,000 (C,D) and 10,000 (E,F) events, the second (h_2) with 10,000–90,000 events (indicated with different symbols). Some symbols for h_2 events are omitted for clarity. Fit curves for Gaussian histograms are shown using solid lines (70,000 and 90,000 h_2 events), long dash lines (40,000 and 60,000 h_2 events), and short dash lines (10,000 and 30,000 h_2 events).

should note that the position of a point in this figure should be interpreted in terms of two QF values representing distances to two respective control populations. As in the case of

DNA ploidy distributions, positions of the symbols were similar when immunofluorescence distributions were populated with large (Fig. 7A), moderate (Fig. 7B), and low (Fig. 7C)
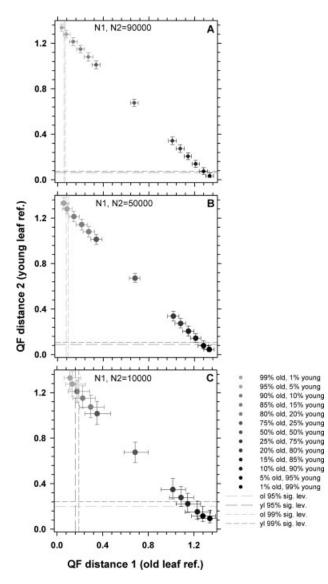
Figure 5. Dependence of critical values (expressed as 95th percentile) of QF distance on standard deviation of histograms constructed from sums of two (**A,D,G**), three (**B,E,H**), and four Gaussian distributions (see Materials and Methods). The first group of compared histograms (h_1) were populated with 90,000 (A,B,**C**), 50,000 (D,E,**F**), and 10,000 (**G,H,I**) events, the second (h_2) with 10,000–90,000 events (indicated with different symbols). Some symbols for h_2 events are omitted for clarity. Fit curves for Gaussian histograms (Fig. 1) are shown using solid lines (70,000 and 90,000 h_2 events), long dash lines (40,000 and 60,000 h_2 events), and short dash lines (10,000 and 30,000 h_2 events).

numbers of events. Likewise, the QF critical values increased with decreasing number of events (compare panels A, B, and C in Fig. 7).

As in the case of DNA ploidy distributions, when the PB-χ^2 statistic was used false negative or false positive bias was obtained (respective fraction from 0.001% to 8.7% at 95% significance level), depending on the histogram shape, numbers

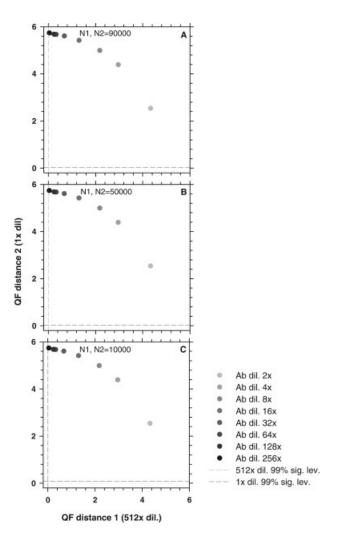of events comprised in the histograms, and number of PB bins. Modified PB-χ^2 exhibited negative bias (the respective fraction from 1.8% to 4.2%). This effect also occurred also when KS was used (the respective fraction was 2.0%–2.4%). One should note that performance of modified PB-χ^2 and KS was better in the simple case of immunofluorescence than in the case of DNA ploidy histograms.

Figure 6. Change in ploidy pattern of *Brassica campestris* leaf tissue with aging. The QF distances of intermediate histograms (see Materials and Methods) were determined with respect to histograms of old (*x* axis) and young (*y* axis) leaf tissue as the reference. The relative contribution of the reference distributions is indicated by shade of the symbols (from 1% young – light gray to 99% young – black). The distributions were populated with 90,000 (**A**), 50,000 (**B**), and 10,000 (**C**) events. Critical QF distance values for old leaf (light gray lines) and young leaf (dark gray lines) were calculated for 95% (long dash) and 99% (short dash) significance levels.

Discussion

Flow cytometry histograms are used to quickly visualize and analyze biological responses, and to provide rapid quality control. Although cytometry histograms are usually compared using only geometric mean, variance, skewness, and kurtosis, a rigorous comparison of cytometry histograms should involve two steps: the first is quantification of (dis)similarity between histograms (a measure of difference), and the second is estimation of the statistical significance of this difference.

Fractions of positive/negative cells (23–25) or cells in a specific phase of the cell cycle (24) are routinely used as difference estimators. The quantification has also been implemented using various methods of histogram subtraction (26), χ^2 (Chi-square) distance (7,8,12), or KS (Kolmogorov–Smirnov) (1,22) metric. The proposed method of cytometry histogram comparison uses quadratic form (QF) and provides a direct estimate of distance, like the two latter methods. However, in contrast to χ^2 or histogram subtraction, The QF measure satisfies the subadditivity condition (triangle inequality); thus it is a true mathematical metric in histogram space. One should

Figure 7. Change in fluorescence intensity distribution of immunostained lymphocytes with concentration of CD45-FITC antibody used for labeling. The QF distances of intermediate histograms (see Materials and Methods) were determined with respect to histograms of populations stained using the highest (512×, *x* axis) and lowest (1×, *y* axis) dilutions of the antibody. The dilution is indicated by shade of the symbols (from 2× - light gray to 256× - black). The distributions were populated with 90,000 (A), 50,000 (B), and 10,000 (C) events. Critical QF distance values for 1× dilution (dark gray lines) and 512× dilution (light gray lines) were calculated for 99% (dashed lines) significance levels. One should note that the lines are close to the axes and may not be clearly visible.

note that the KS distance is the only other histogram difference estimator used in flow cytometry which offers a similar advantage. However, unlike KS, QF employs a ground-distance estimation as an interbin measure (18). Hence, the QF scales accordingly with histogram scale and consequently may be calibrated to account for variable differences between bins (owing to a nonlinear intensity scale, or in the case of comparison of histograms obtained with two different flow cytometers). To our knowledge none of the previously proposed measures offers these advantages.

The estimation of the statistical significance of histogram dissimilarity measures has been approached before using non-parametric tests based on Kolmogorov–Smirnov (KS) (1–3) or χ^2 statistics (7,17). However, in practical applications these tests tend to underestimate the probability with which discrete data sets are unique (8,22). To eliminate this problem a parameterized KS statistic (22) and a probability binning (PB) method (8,12) have been proposed. One should note that the effect of histogram shape on performance of the former method has not been estimated (22). Also, the critical values for the PB-χ^2 statistic (calculated only for Gaussian distributions) proposed by Roederer et al. were not validated for non-Gaussian histograms (8). Furthermore, a Gaussian distribution of the PB-χ^2 critical values was assumed without proof (8). This assumption may not always be fulfilled in practice, as pointed out by Baggerly (12). As a consequence PB-χ^2 may produce a larger-than-expected rate of false positive or negative results depending on the number of events in the compared control and sample histograms. The modified PB-χ^2 proposed by Baggerly (12) exhibited better performance with our data, yet the problem was not completely eliminated. Moreover, PB-χ^2 in either version requires binning of original data. This operation affects both sensitivity and specificity of PB-χ^2 statistics, yet there is no universal procedure (applicable to histograms of any shape) to establish the optimal binning scheme. As a result, PB-χ^2 may produce large numerical errors when histograms have narrow maxima (like those presented in Supplemental Fig 3.). One should note that any flow cytometry histogram contains data which are already assigned to a number of discrete bins owing to digitization of the fluorescence signal. The proposed QF statistic does not require binning and therefore takes advantage of the full dynamic range of cytometric intensity measurements.

It has been demonstrated using the Monte-Carlo method that distributions of QF distances are non-Gaussian. Therefore, distribution percentiles were used to calculate critical values for significance levels. Alternatively, confidence intervals for histogram comparisons could be estimated using the bootstrap method, or by computing an approximation of a QF distribution for a given underlying type of histograms and matrix $\mathbf{A}_j^i$. These methods are not discussed in this report.

It was shown that the critical values (computed for significance levels from 50 to 99%) depended on histogram width and the number of events, but did not depend on histogram shape. Therefore, it has been concluded that one can use the proposed QF statistic to compare flow cytometry histograms of arbitrary shape.

To clearly demonstrate this notion, we applied our metric to compare distributions of DNA content in plant tissue. The histograms representing this type of sample are non-Gaussian and contain multiple maxima, which make dissimilarity measurement especially challenging. Thus, this biological material provides a very demanding test for the distance function. Since the only way to test the analysis method is to employ it with data for which the correct answer is known beforehand, we followed the procedure described by Overton (26), who used computerized mixtures of real fluorescence data from positive and negative controls to test the proposed histogram subtraction methods. Similarly, we used histograms generated in silico, representing various mixtures of controls, and employed QF, PB-χ^2, modified PB-χ^2, and KS statistics to evaluate the differences. The proposed QF statistic, in contrast to both variants of PB-χ^2, does not required histogram binning, and has been proved to operate robustly for non-Gaussian histograms. Therefore, our method performed reliably when distributions contained several narrow peaks as in the case of DNA content histograms of nuclei from leaf tissue of *Brassica campestris*, or one non-Gaussian peak as in the case of lymphocyte immunofluorescence. One should note that PB-χ^2, modified PB-χ^2, and KS did not offer stable and unbiased performance in these conditions.

The proposed statistic was originally constructed using univariate histograms. However, QF may easily be extended to multivariate case. One should note that the standard KS metric is not defined for multivariate data. Extension of PB-χ^2 to multivariate histograms is mathematically straightforward (8,11), but may be technically challenging owing to the necessity of establishing the optimal binning scheme. Furthermore, use of a distance matrix $\mathbf{A}_j^i$ in the definition of QF offers the possibility of incorporating instrument noise models and calibration data directly into the comparison algorithm. In contrast to cross-bin histogram (dis)similarity measures (for instance QF or Wasserstein distance), bin-to-bin measures (such as χ^2 or KS) do not have this advantage. All of the above makes QF distance and QF statistic an ideal tool for future cytometry data-mining systems, where users will be able to search large collections of multivariate cytometry measurements amassed with the help of high-throughput systems for instances that are similar to a specified query. The ability to automatically retrieve similar measurements with arbitrary distributions, even obtained with different instruments, will offer a dramatic improvement in cytometry data management and analysis.

CONCLUSIONS

The objective of this manuscript was to introduce a new quantitative metric that can be used as a tool for histogram dissimilarity evaluation and measurement. Therefore, our goal was to provide evidence that indeed QF is a metric, that it behaves in a stable and predictive manner, and that the results it produces are in agreement with the ground truth. This can be achieved mostly through computer simulations, or use of real but simple and very well characterized biological samples. However, as with any statistical tool used in cytometry, the question can be raised whether the proposed metric always remains relevant for

underlying biological phenomena that cytometry experiments may try to characterize. There is no easy answer to this question. As pointed out by Young (1) and Finch (3) while discussing the applicability of the K-S test for evaluation of differences between cytometry histograms, there is no objective mechanism for demonstrating that statistically significant differences between features of populations measured with the help of flow cytometry indeed imply that these differences are biologically relevant. Additionally, as it has been often pointed out, statistically significant differences are not necessarily practically meaningful when evaluating biological phenomena (3,7,12). In fact, a belief that in the absence of any biological plausibility and prior evidence statistical methods can provide a single number (a *P*-value, a distance measure, etc.) that by itself reflects the probability of occurrence of a certain experimental outcome is just a common misconception (27). Although this limited paper does not attempt an in-depth analysis of inferential reasoning in the life sciences, we would like to stress the importance of proper controls when any type of metric or statistics is used to compare cytometry histograms. Like any other metric, QF operates under the assumption that the property that has been measured is biologically meaningful. If a cytometry practitioner cannot demonstrate that indeed it is the investigated biological differences that lead to changes in outcome of cytometry measurements, then either QF, as well as any other metric, becomes completely useless, or its use is limited to description of instrument drift or stain artifacts. Moreover, when QF is used as a metric in a statistical test one has to remain very careful when interpreting the results (as with any tests performed within the frequentist framework) to avoid overstatement of the evidence against the null hypothesis (28). Finally, we encourage any possible adopters of the QF metric to assess the natural biological variability of their samples and the variability added by their instruments in order to establish a cut-off value such as a minimum important effect size below which the difference is considered negligible, or to design a custom empirical bin-to-bin distance matrix before the metric is utilized to make any predictions regarding the differences between actual samples (29).

Literature Cited

1. Young IT. Proof without prejudice—Use of Kolmogorov-Smirnov Test for analysis of histograms from flow systems and other sources. J Histochem Cytochem 1977;25: 935–941.

2. Bagwell B. A journey through flow cytometric immunofluorescence analyses–Finding accurate and robust algorithms that estimate positive fraction distributions. Clin Immunol Newslett 1996;16:33–37.

3. Finch PD. Substantive difference and the analysis of histograms from very large samples. J Histochem Cytochem 1979;27:800.

4. Huang S, Yeo AA, Li SD. Modification of Kolmogorov-Smirnov test for DNA content data analysis through distribution alignment. Assay Drug Develop Technol 2007;5:663–671.

5. Brescia F, Sarti M. Modification to the Lampariello approach to evaluate reactive oxygen species production by flow cytometry. Cytometry A 2008;73A:175–179.

6. Van Bockstaele F, Janssens A, Piette A, Callewaert F, Pede V, Offner F, Verhasseh B, Philippe J. Kolmogorov-Smirnov statistical test for analysis of ZAP-70 expression in B-CLL, compared with quantitative PCR and IgVH mutation status. Cytometry B 2006;70B:302–308.

7. Cox C, Reeder JE, Robinson RD, Suppes SB, Wheeless LL. Comparison of frequency distributions in flow cytometry. Cytometry 1988;9:291–298.

8. Roederer M, Treister A, Moore W, Herzenberg LA. Probability binning comparison: A metric for quantitating univariate distribution differences. Cytometry 2001;45:37–46.

9. Vives-Rego J, Resina O, Comas J, Loren G, Julia O. Statistical analysis and biological interpretation of the flow cytometric heterogeneity observed in bacterial axenic cultures. J Microbiol Methods 2003;53:43–50.

10. Watson JV. Flow Cytometry Data Analysis. Basic Concepts and Statistics. Cambridge: Cambridge University Press; 1992.

11. Roederer M, Moore W, Treister A, Hardy RR, Herzenberg LA. Probability binning comparison: A metric for quantitating multivariate distribution differences. Cytometry 2001;45:47–55.

12. Baggerly KA. Probability binning and testing agreement between multivariate immunofluorescence histograms: Extending the chi-squared test. Cytometry 2001;45:141–150.

13. Brunelli R, Mich O. Histograms analysis for image retrieval. Pattern Recogn 2001;34:1625–1637.

14. Desealers T, Keysers D, Ney H. Features for image retrieval: An experimental comparison. Image Retrieval 2008;11:77–107.

15. Lee SM, Xin JH, Westland S. Evaluation of image similarity by histogram intersection. Col Res Appl 2005;30:265–274.

16. Khalid MS, Malik MB, Ilyas MU, Sarfaraz MS, Mahmood K. Performance of a similarity measure in grayscale image matching. In: Proceedings of the IEEE Symposium on Emerging Technologies, Islamabad, IEEE, 2005.

17. Rubner Y, Puzicha J, Tomasi C, Buhmann JM. Empirical evaluation of dissimilarity measures for color and texture. Comput Vision Image Understanding 2001;84:25–43.

18. Rubner Y, Tomasi C, Guibas LJ. The Earth Mover's Distance as a metric for image retrieval. Int J Comput Vision 2000;40:99–121.

19. Steinwolf A, Rizzi SA. Non-Gaussian analysis of turbulent boundary layer fluctuating pressure on aircraft skin panels. J Aircraft 2006;43:1662–1675.

20. Winterstein SR. Nonlinear vibration models for extremes and fatigue. J Eng Mech-Asce 1988;114:1772–1790.

21. Hajdera I, Siwinska D, Hasterok R, Maluszynska J. Molecular cytogenetic analysis of genome structure in *Lupinus angustifolius* and *Lupinus cosentinii*. Theor Appl Genet 2003;107:988–996.

22. Lampariello F. On the use of the Kolmogorov-Smirnov statistical test for immunofluorescence histogram comparison. Cytometry 2000;39:179–188.

23. Bigos M, Baumgarth N, Jager GC, Herman OC, Nozaki T, Stovel RT, Parks DR, Herzenberg LA. Nine color eleven parameter immunophenotyping using three laser flow cytometry. Cytometry 1999;36:36–45.

24. Brown M, Wittwer C. Flow cytometry: Principles and clinical applications in hematology. Clin Chem 2000;46:1221–1229.

25. Tung JW, Parks DR, Moore WA, Herzenberg LA, Herzenberg LA. New approaches to fluorescence compensation and visualization of FACS data. Clin Immunol 2004;110:277–283.

26. Overton WR. Modified histogram subtraction technique for analysis of flow-cytometry data. Cytometry 1988;9:619–626.

27. Goodman SN. Toward evidence-based medical statistics, Part 1: The P value fallacy. Ann Int Med 1999;130:995–1004.

28. Goodman SN. A comment on replication, *P*-values and evidence. Stat Med 1992;11:875–879.

29. Parkhurst DF. Statistical significance tests: Equivalence and reverse tests should reduce misinterpretation. Bioscience 2001;51:1051–1057.