Traffic 2002; **3:** 61–65
Munksgaard International Publishers

ISSN 1398-9219

Objective Evaluation of Differences in Protein Subcellular Distribution

Edward J. S. Roques and Robert F. Murphy*

Department of Biological Sciences and Center for Light Microscope Imaging and Biotechnology, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh PA 15213, USA
** Corresponding author: Robert F. Murphy, murphy@cmu.edu*

Research on the cell and molecular biology of proteins involved in membrane traffic often requires determination of the effects of various experimental conditions on the subcellular distributions of those proteins. This is most often accomplished by acquiring fluorescence microscope images and visually comparing these images. While this approach is quite suitable for detecting major changes in distributions, it is not sensitive to small changes and does not permit a quantitative and objective analysis. We therefore describe the application of pattern analysis methods to the comparison of sets of fluorescence microscope images. This approach provides a high throughput and reproducible technique to determine whether image distributions differ within a specified statistical confidence, and is shown to resolve image sets indistinguishable by visual inspection.

Key words: fluorescence microscopy, high-throughput screening, image similarity, pattern analysis, pattern comparison, protein distribution comparison, protein localization, proteomics, subcellular location features

Received 23 July 2001, revised and accepted for publication 23 October 2001

Each protein expressed in a given cell type will localize within that cell according to the properties of the protein, the complement of other proteins expressed in that cell, and the environmental state of the cell. Changes in protein expression or in environmental conditions may result in changes in protein localization. For example, the GLUT4 glucose transporter is redistributed from endomembrane vesicles to the plasma membrane upon exposure to insulin (1–3) and the Menkes protein shifts from the Golgi apparatus to the plasma membrane upon addition of copper (4). In these cases, the change in distribution is clear from visual analysis of fluorescence microscope images. However, in some cases (such as redistribution between two visually similar organelles) changes in protein distribution may not be so evident. For instance, the balance between distribution in endosomes and lysosomes may shift upon various pharmacological treatments (5). In such cases, the task is to determine whether two sets of similar images indeed differ. Since we have previously described automated methods for *classification* of protein subcellular location patterns in *individual* cell images (6–8), we explored the feasibility of extending these methods to perform automated, objective and sensitive *comparison* of *sets* of fluorescence images.

Before beginning this work, we performed extensive literature searches of the PubMed database and the INSPEC database. These searches revealed little prior work that we could build on. For example, searching PubMed with [''protein'' and ''quantitative'' and ''comparison'' and ''localization''] yielded a total of 31 papers. None of these dealt with general methods for comparison of subcellular patterns. Similar searches of INSPEC produced no results.

A potential approach to the problem is to perform a pixel-by-pixel comparison of pairs of images. Since cells show extensive variation in overall shape and in the intracellular distribution of organelles, this approach is not feasible. An alternative approach is to calculate numerical features that describe the subcellular pattern in each image and then compare the features. In this case, the numerical features must be shown to capture the essential characteristic or characteristics of the subcellular patterns of interest. The features can include a set of general descriptors of cell morphology or one or more specific features developed for a particular application.

Limited work using the latter approach has been described previously. For example, the Lippincott-Schwartz group (9–11) quantitated translocation into and out of the Golgi complex in a time series of images by using the image of the target protein at one time point to create a region of interest to measure the fraction of protein in that region at other time points. This method is appropriate in some cases for repeated images of the same field, but not for single images of many cells. We therefore present here a method for quantifying changes in subcellular protein distributions and apply a standard statistical test to determine the significance of those changes.

Results and Discussion

We have previously described methods for the numerical description of protein localization using Zernike moment features and Haralick texture features (6), and more recently have derived an additional set of features specifically for analysis of subcellular location (7,8). These features are invariant to position and rotation, and have been shown to be

capable of distinguishing the major subcellular organelles and structures of cultured cells. The full set of subcellular location features (SLF) consists of 84 features per image (which includes six features derived from a parallel image of the DNA distribution). A collection of images of a labeled cell population can therefore be represented by a matrix where each column contains the numerical values for a given feature and each row contains the features for a different image.

In order to compare two sets of images (e.g. before and after treatment with a drug), our task becomes determining whether these matrices are statistically different. We consider two approaches to this test. The first consists of independently comparing the distributions of each feature for the two sets using univariate t-tests (12). The advantages of this approach are that the results of a test for one feature are not affected by the presence of additional features, and that it gives immediate insight into the source of differences between the two image sets. The disadvantages are that it does not take into account correlations between features, and that it exaggerates the probability of at least one false rejection for a given confidence level (if a test is done for each of 84 features using a 90% confidence level, on average one of them would be expected to fail even if the two image sets are drawn from the same distribution).

Therefore, we have used a multivariate statistical approach known as the Hotelling T^2-test (13). The test results in an F distribution with two degrees of freedom: (i) the number of features and (ii) the combined number of images in both sets minus the number of features. The total number of images used in the comparison therefore must be greater than the number of features. Advantages of this approach are that the T^2 and F-values are a quantitative measure of the similarity between the two distributions, and that a single test of the hypothesis that the two sets are indistinguishable can be made by comparing the F-value to the critical value for a given confidence interval.

We performed two experiments to demonstrate the utility of this method for subcellular protein distributions, using the collection of HeLa cell images captured for our study of classification of subcellular patterns (8). This collection contains 10 classes of subcellular distributions corresponding to all major subcellular structures, and includes classes both distinguishable and indistinguishable to the naked eye. Figure 1 shows an example of two classes that are difficult to distinguish, the localization patterns for giantin and gpp130 (both of which localize to the Golgi apparatus).

Our first experiment was to carry out the T^2-test for all pairs of classes. In order that our method would be as generally useful as possible, we created a 65-feature subset of our 84 SLF features by eliminating six features that require a parallel image of the DNA distribution and 13 Haralick texture features that show some sensitivity to the overall intensity of the images. Table 1 shows F-values indicating the degree of similarity of each pair using the 65 features. As expected, we

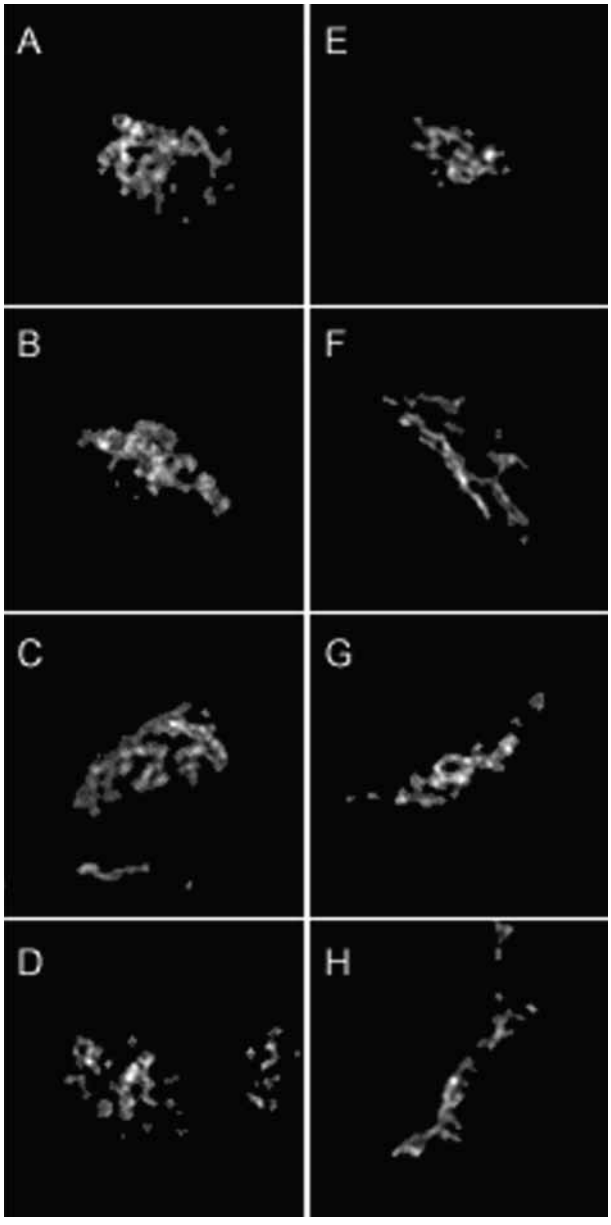

Figure 1: The patterns of two different Golgi proteins are visually indistinguishable by fluorescence microscopy. Shown are the four most typical images (17) of giantin (A–D) and gpp130 (E–H) from the collection of HeLa cell images (8).

observed that each pair of classes was correctly identified as being different. For example, the visually indistinguishable pair giantin and gpp130 gave an F-value of 2.6, which is higher than the critical value of 1.4. The highest values (on the order of 10^2) were found in the comparison of DAPI with other classes, as the DNA distribution is very different from the other subcellular patterns.

To further validate our approach, we wished to establish that two sets of images from the same subcellular protein distribution would be correctly identified as being the same. To

Table 1: T^2-test of similarity between classes of subcellular patterns. Sets of images showing various subcellular patterns were compared. The reagents used to acquire various patterns were 4,6-diamidino-2-pheylindole (DAPI) to label nuclear DNA, phalloidin (Phal.) to label filamentous actin, monoclonal antibodies against unknown antigens specific to the endoplasmic reticulum (ER) and mitochondria (mc151), and monoclonal antibodies against the Golgi proteins giantin and gpp130, the lysosomal membrane protein LAMP2, the nucleolar protein nucleolin, the transferrin receptor (tfr), and the cytoskeletal protein tubulin. The F-values from the T^2-test for the comparison of each class with each other class are shown (the critical values of the F distribution for a 95% confidence level range from 1.42 to 1.45, depending on the total number of images in the comparison). Thus, all classes were considered to be distinguishable from each other with 95% confidence. Note that the lowest F-values were observed for the comparisons of giantin with gpp130 and transferrin receptor with LAMP2 (shown in bold face)

Class	No. of images	DAPI	ER	Giantin	gpp130	LAMP2	mc151	Nucleolin	Phal.	tfr
DAPI	87									
ER	86	90.6								
Giantin	87	138.9	49.9							
gpp130	85	154.1	51.3	**2.6**						
LAMP2	84	92.3	22.6	11.7	11.6					
mc151	73	179.2	11.0	56.1	61.6	17.4				
Nucleolin	73	91.3	60.3	18.7	17.1	20.0	67.0			
Phal.	98	523.5	58.1	374.2	358.2	127.4	17.0	274.2		
tfr	91	101.3	8.6	19.1	17.5	**3.1**	9.0	30.3	26.4	
Tubulin	91	185.5	12.5	97.3	102.4	31.3	8.0	100.5	21.4	6.5

obtain such sets, we randomly divided a single image set into two sets of equal size. We repeated this subdivision and comparison 1000 times, and carried this out for each of the two largest image sets, transferrin receptor and phalloidin. We expected that the sets would be found to be the same 95% of the time and that approximately 50 of the 1000 comparisons would incorrectly identify the sets as being from different distributions. The results shown in Table 2 are in agreement with the expectation.

The possibility remains that reported differences in classes found, especially those of closely related classes such as giantin and gpp130, are due to artifactual differences arising from experimental protocols rather than biologically significant differences in subcellular pattern. As a control for this possibility, we chose to compare sets of images of the same protein acquired under different experimental conditions. To this end, we collected 36 images of cells stained using a rabbit anti-giantin antiserum and 44 images of cells stained using a mouse anti-giantin monoclonal antibody. During the preparation of each set, all conditions were kept as constant as possible. Using the 65-feature subset described above, the two image sets were compared at a statistical confidence level of 95%. The F-value generated was 1.04 relative to a critical value of 2.22, signifying that the two sets are correctly identified as being of the same class, despite the use of different primary and secondary antibodies. The combination of this result with those in Table 2 indicates that our method correctly considers sets that show the same underlying pattern to be the same, and strengthens our conclusion that similar but distinct patterns can be accurately discriminated.

Table 2: T^2-tests of similarity within image classes. Images from a single localization class were randomly divided into two subsets and compared using a set of 65 features. The results shown are compiled for 1000 such random subdivisions. The first degree of freedom was 65 for both tests. The second was 25 for transferrin receptor and 32 for phalloidin. The critical value was calculated from a confidence level of 95% and therefore the expected number of sets failing was 50. The results fall within this expected number of sets failing suggesting that subsets were accurately described as being from the same distribution

	tfr	Phal
Average F	1.08	1.10
Critical value	1.81	1.71
Number of sets failing	32	59

Once two sets are found to be different, it is important to gain insight into the features that contribute to this difference. For this purpose, we have performed univariate t-tests on each feature and recorded the confidence level at which the feature is found to be the same in the two distributions. The features were then ranked according to the confidence levels. Results for this analysis on the giantin and gpp130 sets are shown in Table 3. The major distinction between the two classes arises from differences in the general shape of the protein distribution, the size of the objects, and the distance of the objects from the center of fluorescence. When a similar comparison is done for giantin and phalloidin, 40 of the 49 Zernike moment features and five of the morphological features are all found to differ with greater than 99.99999% confidence.

Our results demonstrate an automated and reproducible method for the comparison of subcellular patterns. We have shown that it is possible to use the F-value derived from a multivariate hypothesis test of two distributions as a numerical indicator of the degree of similarity between images. We have also shown that the system can differentiate

Table 3: Most distinguishing features of a two-class comparison. To isolate the features best able to distinguish giantin and gpp130 images, a univariate t-test was performed on each feature as described in the text

Feature	Confidence level at which feature differs
Eccentricity of the ellipse equivalent to the protein image convex hull	99.99999
Convex hull roundness	99.9999
Measure edge direction homogeneity 1	99.9873
Average number of above threshold pixels per object	99.9873
Average object distance to the center of fluorescence	99.9873
Ratio of largest to smallest object to image center of fluorescence distance	99.9873

between two sets of images previously indistinguishable by visual analysis, thereby surpassing previous subjective techniques. The system can also correctly identify situations in which two patterns showing the same distribution are compared.

With the completion of genome sequencing projects, significant efforts are now being directed to high throughput methods for characterizing the products of the many uncharacterized genes that have been identified. Systematic classification and comparison methods for determining subcellular location for large numbers of proteins can play a role in high throughput characterization projects. In addition, high throughput methods may be used to identify pharmacological agents that have a specific effect on cellular processes, such as protein subcellular location. The objectivity, sensitivity and automation afforded by the approach described here are expected to be of significant value in high throughput screening protocols. Although our results were obtained using manual cropping of images for the purpose of segmenting cells, a number of approaches to fully automated segmentation have been developed previously with favorable results, and these approaches can easily be combined with the methods described here to provide a fully automated system.

An alternative method for sensitive comparison of the distributions of two proteins is the simultaneous use of two or more fluorophores. This approach has been used to distinguish cis and trans Golgi markers (14). However, it has the disadvantage that it requires both proteins to be visualized in the same cells (which is often difficult if both proteins are 'tagged' the same way, such as with GFP fusions or with monoclonal antibodies derived from the same source). It is also unsuited to comparing distributions from two populations of cells that have been treated differently (e.g. with and without a drug). The method we describe is thus much easier to apply and more generally applicable but is not intended to replace two-color analysis for some applications.

A web service (SImEC, for Statistical Image Experiment Comparator) implementing the method described here is available to compare sets of uploaded images (http://murphylab.web.cmu.edu/Services/SImEC).

Materials and Methods

Computational methods

The system described here was implemented in Matlab (The Mathworks, Natick, MA, USA), using the Image Processing and Statistics toolboxes along with previously developed C++ functions for feature calculation (8).

Immunofluorescence microscopy

HeLa cells were grown for 1 d in Dulbecco's modified Eagle's medium (Hyclone, Logan, VT, USA) containing 10% calf serum and 1% penicillin/streptomycin mix (v/v) (Intergen Co., Purchase, NY, USA) on 19-mm cover slips coated with 0.1% (w/v) type I collagen in 0.01 HCl. Cells were washed twice with phosphate-buffered saline (PBS: 140 mM NaCl, 2.6 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 0.9 mM $CaCl_2$, 0.5 mM $MgCl_2$, pH 7.4), fixed for 15 min with 3.7% paraformaldehyde in PBS, and washed twice in PBS. They were then permeabilized for 10 min with 0.1% Triton X-100 in wash buffer (WB: 0.5% bovine serum albumin and 0.15% glycine in CSB, 137 mM NaCl, 5 mM KCl, 1.1 mM Na_2HPO_4, 0.4 mM KH_2PO_4, 4 mM $NaHCO_3$, 2 mM $MgCl_2$. $6H_2O$, 2 mM EGTA, 0.1% glucose, 5 mM Pipes, pH 6.1), washed twice with WB, and blocked overnight at 4°C with goat serum diluted 1 : 20 in WB.

Cells were washed twice with WB and incubated for 60 min with either of two anti-giantin antibodies. The first was a mouse monoclonal antibody directed against the central portion of the rod domain of giantin (clone 133) (15), prepared at a 1 : 200 dilution in WB. The second was a rabbit anti-giantin antiserum directed against the N-terminal portion of giantin (16) prepared at a 1 : 2000 dilution in WB. Cells were washed three times with WB and incubated for 60 min with an appropriate secondary antiserum, either a 1 : 400 dilution of goat anti-mouse IgG (H + L) antiserum conjugated to Alexa Fluor 488 (Molecular Probes, Inc., Eugene, OR, USA) or a 1 : 200 dilution of goat anti-rabbit IgG antiserum conjugated to fluorescein isothiocyanate isomer I (Sigma). After washing three times each with WB and CSB, cells were incubated for 15 min in 100 μg/ml RNAse A (Sigma) in CSB, and washed three times in CSB. Cells were incubated for 15 min in 25 μg/ml FluoroLink Cy5 (Amersham Pharmacia Biotech, Piscataway, NJ, USA) in CSB to label total cell protein, and washed three times each in WB and CSB. Finally, total DNA was stained with 5 μg/ml propidium iodide (Molecular Probes, Inc.) in CSB.

Coverslips were mounted on microscope slides in gelvatol [60 ml of 10 mM Tris, 15 g Airvol 205 (Air Products, Allentown, PA), 30 ml glycerol, 1 g n-propyl gallate]. Three color images (1024 × 1024 pixels, 9.8 mm^2) were acquired using a 100×, 1.4 numerical aperture objective lens mounted on a Leica TCS NT confocal laser scanning microscope (each pixel corresponds to a square region of the sample 98 nm on a side). The DNA and total protein images were used to identify individual cells but features were calculated using only the Alexa Fluor 488 or fluorescein images.

Acknowledgments

We thank Meel Velliste for helpful discussions, Simon Watkins for confocal microscopy expertise, and Adam Linstedt for providing antibodies and helpful discussions. The research described here was supported in part by grant 99–295 from the Rockefeller Brothers Fund Charles E. Culpeper Biomedical Pilot Initiative, by NSF Science and Technology Center grant MCB-8920118, and by NIH grant R33 CA83219.

References

1. Martin S, Millar CA, Lyttle CT, Meerloo T, Marsh BJ, Gould GW, James DE. Effects of insulin on intracellular GLUT4 vesicles in adipocytes: evidence for a secretory mode of regulation. J Cell Sci 2000;113:3427–3438.
2. Chinni SR, Shisheva A. Arrest of endosome acidification by bafilomycin A1 mimics insulin action on GLUT4 translocation in 3T3-L1 adipocytes. Biochem J 1999;339:599–606.
3. Holman GD, Cushman SW. Subcellular localization and trafficking of the GLUT4 glucose transporter isoform in insulin-responsive cells. Bioessays 1994;16:753–759.
4. La Fontaine S, Firth SD, Lockhart PJ, Brooks H, Parton RG, Camakaris J, Mercer JF. Functional analysis and intracellular localization of the human menkes protein (MNK) stably expressed from a cDNA construct in Chinese hamster ovary cells (CHO-K1). Hum Mol Genet 1998;7:1293–1300.
5. Chen CS, Rosenwald AG, Pagano RE. Ceramide as a modulator of endocytosis. J Biol Chem 1995;270:13291–13297.
6. Boland MV, Markey MK, Murphy RF. Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. Cytometry 1998;33:366–375.
7. Murphy RF, Boland MV, Velliste M. Towards a systematics for protein subcellular location: quantitative description protein localization patterns automated analysis of fluorescence microscopy images. Proc Int Conf Intell Syst Mol Biol 2000;8:251–259.
8. Boland MV, Murphy RF. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. Bioinformatics 2001;17:in press.
9. Hirschberg K, Miller CM, Ellenberg J, Presley JF, Siggia ED, Phair RD, Lippincott-Schwartz J. Kinetic analysis of secretory protein traffic and characterization of golgi to plasma membrane transport intermediates in living cells. J Cell Biol 1998;143:1485–1503.
10. Presley JF, Cole NB, Schroer TA, Hirschberg K, Zaal KJ, Lippincott-Schwartz J. ER-to-Golgi transport visualized in living cells. Nature 1997;389:81–85.
11. Hirschberg K, Lippincott-Schwartz J. Secretory pathway kinetics and *in vivo* analysis of protein traffic from the Golgi complex to the cell surface. FASEB J 1999;13(Suppl 2):S251–S256.
12. Prescott P. Student's t-test. In: Kotz S, Johnson NL, Read CB, eds. Encyclopedia of Statistical Sciences. New York: Wiley; 1981: p. 43–49.
13. Pillai KCS. Hotelling's T^2. In: Kotz S, Johnson NL, Read CB, eds. Encyclopedia of Statistical Sciences. New York: Wiley; 1981: p. 669–673.
14. Shima D, Haldar K, Pepperkok R, Watson R, Warren G. Partitioning of the Golgi apparatus during mitosis in living HeLa cells. J Cell Biol 1997;137:1211–1228.
15. Linstedt AD, Hauri HP. Giantin, a novel conserved Golgi membrane protein containing a cytoplasmic domain of at least 350 kDa. Mol Biol Cell 1993;4:679–693.
16. Puthenveedu M, Linstedt A. Evidence that Golgi structure depends on a p115 activity that is independent of the vesicle tether components giantin and GM130. J Cell Biol 2001;155: 227–238.
17. Markey MK, Boland MV, Murphy RF. Towards objective selection of representative microscope images. Biophys J 1999;76:2230–2237.