

# Determining the distribution of probes between different subcellular locations through automated unmixing of subcellular patterns

Tao Peng<sup>a,1</sup>, Ghislain M. C. Bonamy<sup>b,c,1</sup>, Estelle Glory-Afshar<sup>a</sup>, Daniel R. Rines<sup>b</sup>, Sumit K. Chanda<sup>b,d</sup>, and Robert F. Murphy<sup>a,e,f,2</sup>

<sup>a</sup>Center for Bioimage Informatics and Department of Biomedical Engineering and <sup>e</sup>Lane Center for Computational Biology and Departments of Biological Sciences and Machine Learning, Carnegie Mellon University, Pittsburgh, PA 15213; <sup>b</sup>Genomics Institute of the Novartis Research Foundation, San Diego, CA 92121; <sup>c</sup>Hudson-Alpha Institute for Biotechnology, Huntsville, AL 35806; <sup>d</sup>Program of Inflammatory Disease Research, Burnham Institute for Medical Research, San Diego, CA 92037; and <sup>f</sup>Freiburg Institute for Advanced Studies, Albert-Ludwigs-University Freiburg, 79104 Freiburg, Germany

Edited by Ronald D. Vale, University of California, San Francisco, CA, and approved December 23, 2009 (received for review October 30, 2009)

Many proteins or other biological macromolecules are localized to more than one subcellular structure. The fraction of a protein in different cellular compartments is often measured by colocalization with organelle-specific fluorescent markers, requiring availability of fluorescent probes for each compartment and acquisition of images for each in conjunction with the macromolecule of interest. Alternatively, tailored algorithms allow finding particular regions in images and quantifying the amount of fluorescence they contain. Unfortunately, this approach requires extensive hand-tuning of algorithms and is often cell type-dependent. Here we describe a machine-learning approach for estimating the amount of fluorescent signal in different subcellular compartments without hand tuning, requiring only the acquisition of separate training images of markers for each compartment. In testing on images of cells stained with mixtures of probes for different organelles, we achieved a 93% correlation between estimated and expected amounts of probes in each compartment. We also demonstrated that the method can be used to quantify drug-dependent protein translocations. The method enables automated and unbiased determination of the distributions of protein across cellular compartments, and will significantly improve imaging-based high-throughput assays and facilitate proteome-scale localization efforts.

automated microscopy | fluorescence microscopy | location proteomics | pattern recognition | high content analysis

Eukaryotic cells are organized into a number of distinct subcellular compartments and structures that play critical roles in cellular functions. Therefore, protein localization is a tightly regulated process whose failure may lead to severe pathologies (1). Nuclear translocation is well-known example of a regulated subcellular compartmentalization event, and it has been the subject of many studies uncovering its association to human pathologies, such as cancer (1–4). Because the nucleus is a large and easily distinguishable structure, such translocation can readily be measured in an automated fashion. Unfortunately, detecting and quantifying the amount of macromolecules in smaller organelles are much harder tasks to perform using traditional techniques, and they are not readily automated. The problem is made more difficult by the fact that proteins (and other macromolecules) are often found in more than one subcellular structure.

The traditional approach to determining the amount of a macromolecule in specific compartments is measuring colocalization of that molecule with organelle-specific markers labeled with a different fluorophore. This approach is often used for one or a few proteins whose possible locations are known a priori, and recent work has described methods for automatically measuring the fraction of colocalization (5, 6); however, it is difficult to apply on a proteome-wide basis because it requires

the availability of probes for all possible compartments and the collection of separate images of each protein in combination with each marker. An exciting automated approach that builds on the basic colocalization method is the MELK technology, which uses robotics to carry out successive rounds of staining for different macromolecules on a given cell sample or tissue (7). However, this approach is restricted to fixed samples and cannot be used for analysis of dynamic pattern changes in living cells. Hand-tuned algorithms for distinguishing particular subcellular regions are also widely used as an alternative to colocalization in high-content screening (8), but they typically are only able to distinguish major cellular regions and are not easily transferred to the analysis of other regions or cell types.

Beginning with the demonstration that automated recognition of subcellular patterns was feasible (9, 10), our group and others have created systems that are able to classify all major subcellular location patterns, and to do so with a higher accuracy than visual analysis (11–14). Automated systems can also learn what subcellular patterns are present in large collections of images without prior knowledge of the possible patterns (13, 15). However, such pattern-clustering approaches have two major limitations. First, they treat each unique combination of major, fundamental patterns as a new pattern because they cannot readily identify the fundamental patterns of which it is composed. Second, they are not designed to handle cases where the fraction of mixing between two patterns can vary continuously, because such continua are either considered as one large pattern or arbitrarily divided into subpatterns.

Thus, we sought to develop tools to quantify the amount of fluorescence in each compartment for images containing a mixture of fundamental patterns, assuming that sets of images containing each fundamental pattern are available. We have previously proposed an object-based approach (16) to this problem that consisted of two learning stages: learning what object types are present in the fundamental patterns, and learning how much fluorescence is present in each object type in each pattern. The fraction of fluorescence in each fundamental pattern for a mixed image was then estimated by determining the mixture coefficients that were most likely to have given rise to that image. This method was tested on synthetic images created

Author contributions: G.M.C.B., D.R.R., S.K.C., and R.F.M. designed research; T.P., G.M.C.B., and E.G.-A. performed research; G.M.C.B. contributed new reagents/analytic tools; T.P., E.G.-A., and R.F.M. analyzed data; and T.P., G.M.C.B., E.G.-A., and R.F.M. wrote the paper.

The authors declare no conflict of interest.

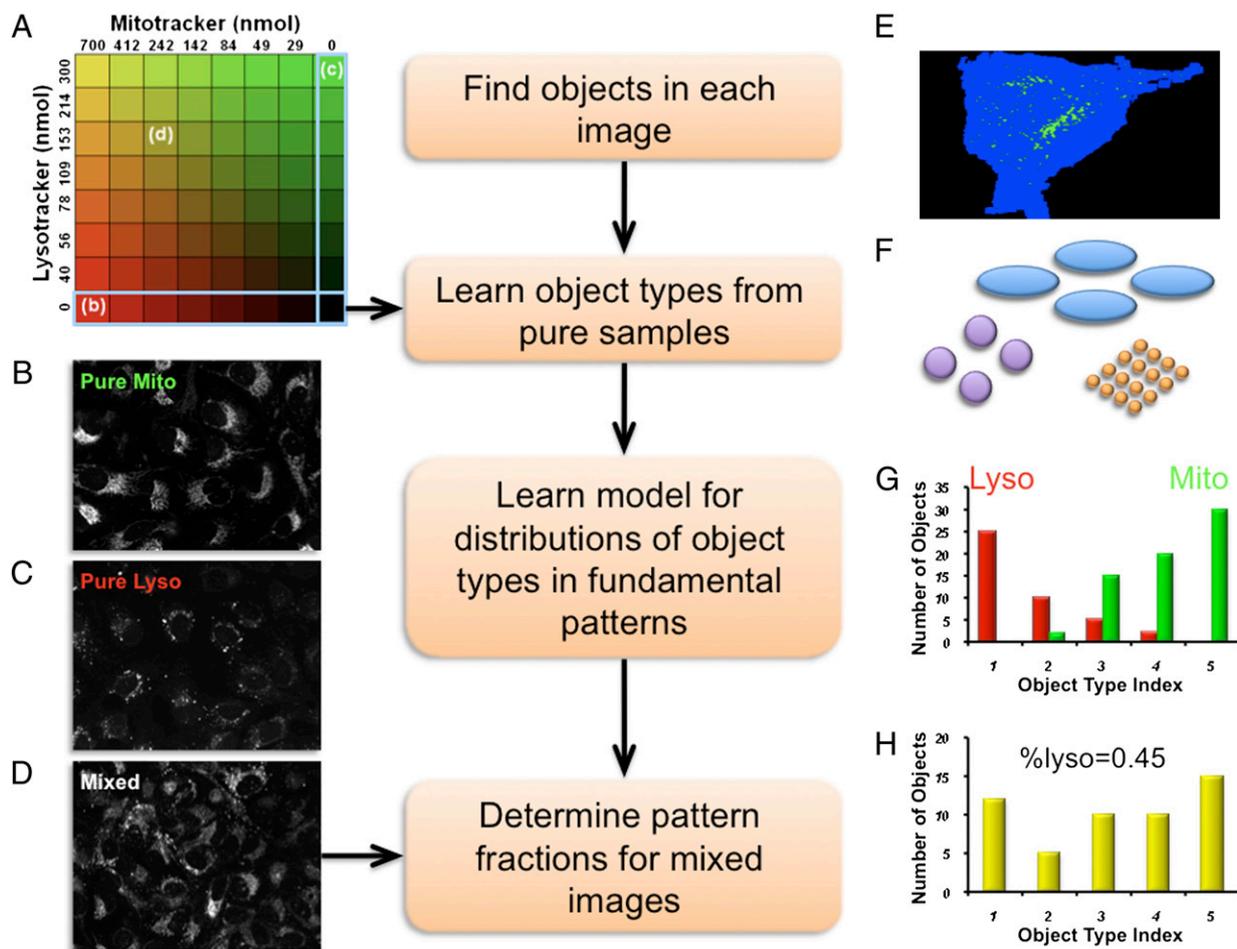
This article is a PNAS Direct Submission.

<sup>1</sup>T.P. and G.M.C.B. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. E-mail: murphy@cmu.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0912090107/DCSupplemental](http://www.pnas.org/cgi/content/full/0912090107/DCSupplemental).





**Fig. 1.** Subcellular pattern unmixing approach. (A) The starting point is a collection of images (typically from a multiwell plate) in which various concentrations of two probes are present (the concentrations of the Mitotracker and LysoTracker probes are shown by increasing intensity of red and green, respectively). Example images are shown for wells containing just Mitotracker (B), just LysoTracker (C), or a mixture of the two probes (D). The steps in the analysis process are shown: finding objects (E), learning object types (illustrated schematically as objects with different sizes and shapes) (F), learning the object type distributions for the two fundamental patterns (G), and unmixing a mixed object type distribution (H).

the testing images is not. Each protein object in a testing image was therefore assigned to the cluster whose center was closest to it in the feature space. The frequency and the total fluorescence of all objects belonging to the same type were calculated for each test image.

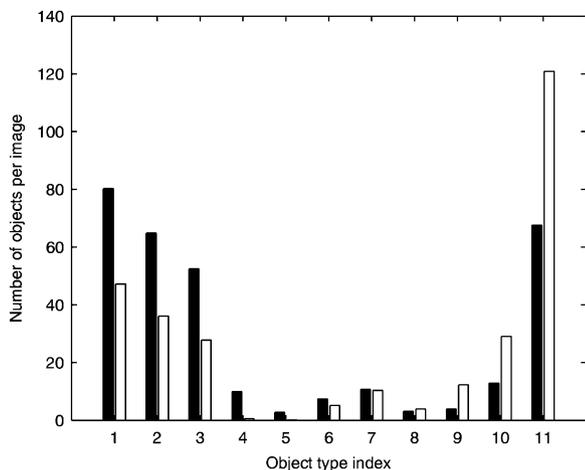
The average fraction of mitochondrial and lysosomal patterns in each well was then estimated by three different approaches using both the object types and object features (see *Materials and Methods*). Two of these methods use the number of objects of each type to estimate mixing fractions. The third uses the amount of fluorescence in each type, which depends on the assumption that this amount is linearly dependent upon the concentration of each probe (Fig. S2).

The results for the three methods are compared with those expected from the relative probe concentrations in Fig. 3. The correlation coefficients between estimated and expected fractions are 0.73, 0.77, and 0.83. An example of the correlation between estimated and actual concentrations is shown in Fig. S3. Analysis of the effect of removing nondiscriminative object types is shown in Fig. S4.

**Removing Outliers.** The unmixing method can be applied without any restrictions to any test images. However, if the mixed patterns in test images contain additional fundamental patterns not present in the training sets, the estimates of mixing fractions will

be incorrect. Our solution consists of two steps. The first step is to exclude objects in test images that do not appear to belong to any of the object types found in the training images (outlier objects). The second step is to flag as outlier patterns those test images that during unmixing have large fitting error (i.e., rejecting images that cannot be successfully decomposed into the specified fundamental patterns). Because we cannot know a priori what kinds of new objects might be encountered, both steps use hypothesis-based tests to find thresholds that retain high accuracy for unmixing the training patterns. To test this approach, we used images in which either the nucleus or the endoplasmic reticulum (ER) was marked. Our results show that this methodology can completely remove nuclear objects during the first level outlier detection (Fig. 4A). For the more difficult case of ER staining, the second level detection recognizes most of the ER-containing images as outlier patterns but retains high accuracy of fundamental pattern unmixing (Fig. 4B).

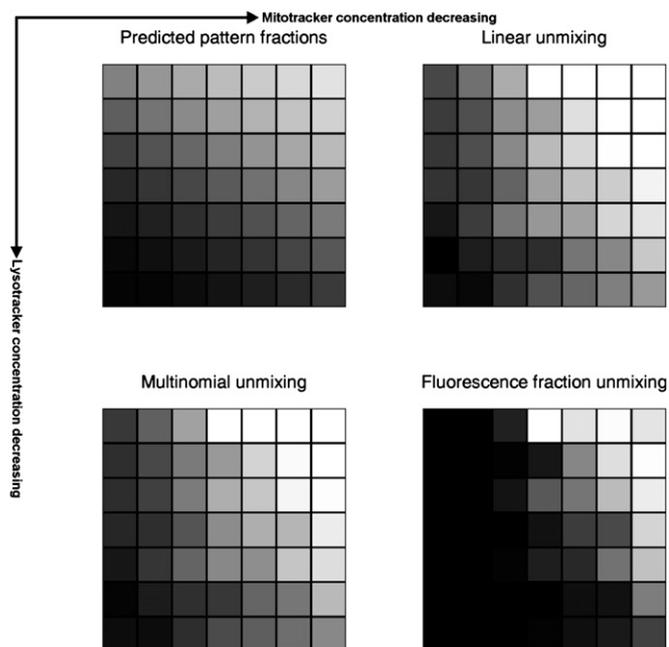
**Application Example: Unmixing Drug-Treated Cell Images.** To further test the usefulness of our unmixing algorithms, we tested whether these tools could correctly estimate the accumulation of the microtubule associated protein LC3 into autophagosomes upon BAF treatment. BAF, an inhibitor of the vacuolar ATPase (17, 18), can inhibit autophagy and promotes autophagosome accumulation by preventing vesicular acidification (19). In this experiment, an



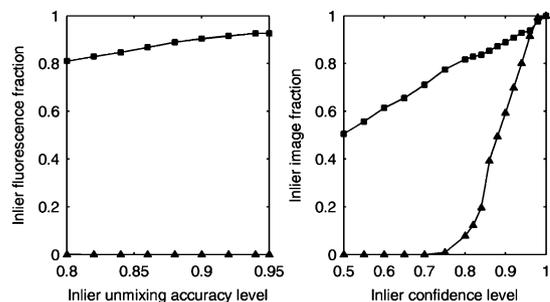
**Fig. 2.** Distribution of object types within fundamental patterns. The average number of objects of each type is shown for the combination of all images of U2OS cells stained with either Mitotracker (black) or Lysotracker (white). The object types are sorted according to the difference between the numbers of objects in the two patterns. Thus, the lowest numbered object types are primarily found in Mitotracker-stained cells, while the highest numbered object types are primarily found in Lysotracker-stained cells. The model power is 0.448 when trained with object frequency distributions and 0.654 when trained with fluorescence fraction distributions.

RT112 cell line stably expressing eGFP-LC3 was treated with various amounts of BAF for 4 h before being fixed and imaged.

To establish the fundamental patterns, the model was trained on images of cells treated with the maximum concentration of drug and of cells that did not undergo any treatment. Unmixing



**Fig. 3.** Expected and estimated pattern fraction for three unmixing methods for the U2OS dataset. The balance between lysosomal and mitochondrial pattern (either expected or estimated) is represented as the fraction of the total pattern (black, 100% mitochondrial; white, 100% lysosomal). For the expected fraction, this is estimated as linearly proportional to the ratio of the relative concentration of the mitochondrial probe to the sum of the relative concentration of the lysosomal and mitochondrial probes (where relative concentration is defined as fraction of the maximum subsaturating concentration).

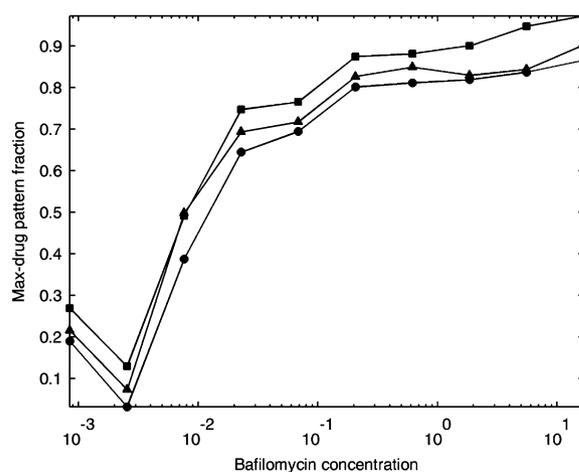


**Fig. 4.** Effectiveness of outlier removal methods. (A) Nuclear images were used as outliers. Unmixing accuracies for both inliers (squares, mitochondrial and lysosomal objects) and outlier objects (triangles) with first-level outlier exclusion were approximated by cross validation under different chosen accuracy levels for the U2OS dataset. Nuclear fluorescence was totally removed at all accuracy levels. (B) ER pattern images were used as outliers. Average outlier recognition testing-accuracies for both inliers (squares, mitochondrial and lysosomal images) and outlier images (triangles) with second-level outlier exclusion were approximated by cross validation under different chosen accuracy levels for the BEAS2B dataset. The best separation is obtained using a 75–80% inlier confidence level.

using different methods resulted in a sigmoid-like transition between the two patterns as a function of the drug concentration (Fig. 5). This example illustrates how our unmixing analysis can be applied to provide a quantitative analysis of protein dynamics.

## Discussion

Our group has previously described approaches to cluster proteins based on their subcellular distribution (13, 15). A logical extension of this work is to create tools to estimate the distribution of fluorescently labeled macromolecules between distinct compartments, and we have previously demonstrated such approaches provide good results for synthetic images (16). Here we show that this approach works well on real images obtained from mixed patterns and is suitable for high-throughput microscopy, technology that would arguably benefit the most from such a strategy. Our test mimicked the case of a tagged protein whose



**Fig. 5.** Application of pattern unmixing to drug effects. Cells were treated with various concentrations of Bafilomycin a1 (BAF) and images from samples receiving the highest dose and images from untreated cells were used to train an unmixing model. The fraction of drug treated pattern as a function of concentration of drug was estimated using linear unmixing (squares), multinomial unmixing (circles), and fluorescence fraction unmixing (triangles). eGFP-LC3 showed a gradual relocation between the two patterns as a function of BAF concentration.

distribution varies between two organelles. Because we controlled the amount of both dyes applied to a given cell sample, it was easy to verify whether or not our predictions about the proportion of mitochondrial or lysosomal labeling were accurate. The successful results described here validate the effectiveness of the two-stage, object-based unmixing method on real image data. They also validate our assumption that new object types that might arise do not drastically inhibit unmixing accuracy.

The tool we have described requires only a set of images for each of the pure patterns (for varying levels of expression, if desired) and a set of images for mixed patterns acquired under the same conditions. Incorporation of outlier tests and accuracy estimates makes the approach robust to unanticipated phenotypes. It is also important to note that the fluorophores used for training and testing need not correspond. For example, a system trained on GTP-tagged proteins could be used to unmix an image of an RFP-tagged protein (assuming that the tags do not alter the protein localization). However, because most cell types show different morphologies, it is unlikely that a system trained for one type could be applicable to others.

The success of the experiments described here should provide the capacity to better describe what may be complex effects of drugs or disease on protein location. The tool offers a previously unexplored method to determine a precise and objective subcellular distribution of gene products for various physiological contexts and genetic backgrounds. This approach can also aid large-scale projects, such as proteome-wide localization studies, because it has been tested for images acquired using automated microscopy.

## Materials and Methods

**Tissue Culture.** U2OS cells were grown in D-MEM (Invitrogen), containing 10% FBS (HyClone). The BEAS2B cell line was grown in a F12/D-MEM[1:1] (Invitrogen) containing 10% FBS (HyClone). All media were supplemented with 100  $\mu\text{g}/\text{mL}$  of Penicillin-streptomycin and L-Glutamine. A total of 7,000 cells per well were plated in 384-well plates in Phenol-red free OptiMem (Gibco) containing 2% serum. After 24 h, Phenol red-free OptiMem containing various concentrations (see Fig. 1) of either Mitotracker green or LysoTracker green (Molecular Probes), as well as 2.5  $\mu\text{g}/\text{mL}$  Hoechst 33342 (Molecular Probes), was added. In some cases, ERTracker green (Molecular Probes) was added to test the outlier removal algorithm. A total of six replicates of an  $8 \times 8$  matrix of various concentrations of Mitotracker and LysoTracker were dispensed in each plate. After a 30-min incubation, the living cells were imaged by high throughput microscopy.

For experiments analyzing drug effects, RT112 cells were plated in a 1,536-well plate at 500 cells per well and grown for 24 h before being treated with various amounts of BAF. After 4 h, cells were fixed with  $1 \times$  Mirskys fixative and stained with Hoechst as above.

**High-Content Imaging.** Images were acquired using an Opera automated microscope equipped with Nipkow confocal spinning disks (Perkin-Elmer). Cells were kept viable by a built-in incubation chamber. The Hoechst stain, which marks the nuclei, was detected using 405-nm laser excitation and a 450/50-nm emission filter, while Mitotracker, LysoTracker, ERTracker, and eGFP (enhanced GFP) were detected with 488-nm laser excitation and a 535/50 emission filter. To minimize fluorescence cross-talk, the two channels were collected sequentially. To provide a sufficient spatial resolution for the analysis, images were collected using a  $40 \times$ , 0.9 NA water immersion objective. For each well of the 384-well plate, 35 adjacent fields with 1.5% overlap organized in a  $7 \times 5$  rectangle were collected. Each field usually contained about 20 cells; a small number of out-of-focus images in some wells were automatically discarded.

**Image Analysis. Preprocessing.** Images containing no nuclei and out-of-focus images were removed by thresholding on total Hoechst fluorescence. Shading and skew corrections were performed on remaining images to compensate for nonhomogeneous illumination and differences in registration between channels. Background fluorescence was removed by subtracting the most common pixel value from each pixel.

**Object detection.** An automated threshold method (20) was used to distinguish probe-containing from nonprobe-containing pixels. In the resulting binary

images, each set of connected above-threshold pixels (an object) was identified. Objects containing fewer than 5 pixels were ignored. The same approach was applied on the DNA channel to identify DNA objects.

**Object feature calculation.** To describe the properties of each object, a set of subcellular object features, SOF1, was calculated as described previously (16). This set is composed of nine features based on morphological properties of the object and two features describing the spatial relationship between objects and the cell nucleus. However, because in the experiments described here images were not segmented into single cells, feature SOF1.2 was replaced by the average distance between each object and the nearest nucleus. All features were normalized to zero mean and unit standard deviation calculated using the training data (Z-scores).

**Object-type learning.** The features for all objects from the singly-stained samples were clustered using the NetLab *k*-means function. The quality of the clustering was assessed using the Akaike Information Criterion as described previously (16). For test images, objects were assigned to the cluster whose center was the smallest Euclidean distance from it.

**Linear unmixing.** Once the *k*-object types are defined, each image can be represented as a vector  $y = (y_1, \dots, y_k)$  of the frequency of each object type in that image. We define  $u$  as the total number of fundamental patterns. For  $u = 2$ , a mixture of pattern 1 (lysosomal) with  $n_1$  objects of a specific object type and pattern 2 (mitochondrial) with  $n_2$  objects of the same object type is assumed to generate a mixed pattern with  $n_1 + n_2$  objects of this type.

We assume that mixed pattern object frequencies are linear combinations of fundamental pattern object frequencies  $f_p$  as follows:

$$\begin{cases} y = \sum_{p=1}^u \alpha_p f_p \\ \sum_{p=1}^u \alpha_p = 1 \end{cases}$$

where  $\alpha_p$  represents the proportion of fundamental pattern  $p$  in the composition of the mixture. Therefore, a mixed pattern can be represented by a vector of coefficients  $a = (\alpha_1, \dots, \alpha_u)$ , containing the fraction of fundamental patterns of which it is composed. Unmixing the mixture pattern consists of solving the linear equation above. Because we have *k* equations for all object types and only two fundamental patterns ( $u = 2$ ), a reasonable solution is to minimize the squared error  $SE = \sum_{i=1}^k (\hat{y}_i - y_i)^2$  under constraint  $\sum_{i=1}^k \alpha_p = 1$ . Because the contribution of any fundamental pattern cannot be negative, nonnegative constraints  $\alpha_p \geq 0$  are included. The solution was found using quadratic programming methods (the Matlab `lsqlin` function).

**Multinomial unmixing.** An alternative method for unmixing is based on the fact that the number of objects of each type varies between cells even within the same pattern (e.g., the number of small lysosomes can vary from cell to cell). We can reasonably assume that if we learn the distribution of the number of objects per cell of a given type for a given fundamental pattern that it will also apply to the distribution of that object type in a mixed sample. In a multinomial distribution, each object belongs in exactly one of the *k* possible object types with the probabilities  $(\theta_1, \dots, \theta_k)$  (so that  $\theta_1 \geq \theta_k$  and  $\sum_{i=1}^k \theta_i = 1$ ). Therefore, each fundamental pattern is represented by a multinomial distribution  $(\theta_1^p, \dots, \theta_k^p)$  where  $\theta_i^p$  is the probability that an object from pattern  $p$  belongs to the object type  $i$ . This can be estimated by the maximum likelihood estimator of a multinomial distribution:

$$\hat{\theta}_i^p = \frac{n_i^p}{\sum_{i=1}^k n_i^p}$$

where  $n_i^p$  corresponds to the number of objects of pattern  $p$  which are of type  $i$ . Mixed patterns are represented by a multinomial distribution composed of a linear combination of fundamental pattern distribution parameters,

$$\theta = \left( \sum_{p=1}^u \alpha_p \theta_1^p, \dots, \sum_{p=1}^u \alpha_p \theta_k^p \right)$$

To reach a distribution which best fits the data of mixed pattern conditions, we adjust the coefficients  $\alpha_p$  to maximize the likelihood of the object frequency for all *k* types:

$$\hat{\alpha} = \max_{\alpha} \prod_{i=1}^k \left( \sum_{p=1}^u \alpha_p \theta_i^p \right)^{\sum_{p=1}^u n_i^p}$$

An equivalent problem would be to maximize the log likelihood. We have previously deduced closed forms of the first-order derivative and the Hessian matrix, and proved concaveness of the log likelihood function (16). Therefore, Newton's method was adopted to solve the optimization problem.

**Fluorescence fraction unmixing.** In addition to using the number of objects of each object type to estimate pattern fractions, we can also use the amount of fluorescence in each object type. To do this, we first find the average fraction of fluorescence  $f_i^p$  within each object type  $i$  for each fundamental pattern  $p$ . We also determine a constant  $L_k$  for each probe that relates the concentration of that probe to the expected amount of total fluorescence in all objects of the pattern  $k$  labeled by that probe. We combine these to calculate  $\lambda_k^p$  as the amount of fluorescence within each object type that is expected per unit of probe added (assuming that fluorescence is roughly linear over the range of probe concentrations added). The fluorescence expected in each object type  $F_k$  is then given by

$$\begin{pmatrix} F_1 \\ \vdots \\ F_k \end{pmatrix} = \sum_{p=1}^u C_p \begin{pmatrix} \lambda_1^p \\ \vdots \\ \lambda_k^p \end{pmatrix}$$

where  $C_p$  is the concentration of probe that labels pattern  $p$ . We used the pseudoinverse approach to estimate this amount.

**Outlier detection.** To address the possibility that a particular pattern being unmixed is not a mixture of the fundamental patterns used during training (that is, that it might contain other patterns as well), we developed a two-level outlier detection method. The unmixing results were expressed as  $(\alpha_0, \alpha_1, \dots, \alpha_u)^T$ , where  $\alpha_0$  is the fluorescent fraction of any unrecognized pattern (outliers) and  $u$  is the total number of fundamental patterns. Both levels use statistical hypothesis tests to determine outliers. First, we used a  $\chi^2$  test to remove outlier objects that were not similar to any of the object types learned from the fundamental pattern images. The  $\chi^2$  statistic is defined as  $x^2 = \sum_{i=1}^m x_i^2$ , where  $x_j$  is the  $j^{\text{th}}$  feature of an object and  $m$  is the total number of features. Statistics of test object  $x_j^2$  were tested under every  $\chi^2$  distribution learned from each object type in the training images to see if it was from that distribution. An object was considered an outlier if it was rejected by all tests at a specified confidence level  $\alpha$ . Because a proper value for  $\alpha$  is hard to determine a priori, we chose it by a linear search using unmixing of the fundamental patterns. The fundamental pattern images were split into training and test sets and the accuracy was reported as the

fraction of objects that were associated with the correct pattern. For various  $\alpha$ , we used cross validation to get averaged accuracies. Accuracy improves with decreasing  $\alpha$  cut-offs, in other words, with a stricter criterion, more objects are excluded as outliers. We chose an arbitrary acceptable accuracy level and its associated level  $\alpha$  to remove outlier objects in testing images.

Similarly, we performed a hypothesis test to exclude mixed patterns that had large fitting errors when decomposed into fractional combination of fundamental pattern fluorescence fractions. This fitting error statistics was defined as:

$$E = \left\| \mathbf{F} - \sum_{p=1}^u \alpha_p \Delta^p \right\|, \mathbf{F} = (F_1, \dots, F_k)^T \text{ and } \Delta^p = (\lambda_1^p, \dots, \lambda_k^p)$$

Statistics of a test-pattern fluorescence fraction  $E_T$  was compared with the empirical distribution of the fitting error. Pattern was rejected to be a mixture of fundamental patterns if  $E_T$  was beyond a certain threshold  $T_2$  ( $E_T > T_2$ ). A large  $T_2$  value tolerates more possible real mixture patterns but also risks accepting more unknown patterns. We defined the accuracy of this level as the fraction of training images not rejected. We next learned the empirical distribution of the fitting error on the training set. We then chose  $T_2$  corresponding to an appropriate accuracy level.

**Unmixing result evaluation.** The significance of the unmixing result of a mixture pattern can be evaluated by its distinctiveness to the composing fundamental patterns. We defined the model power as a quantitative evaluation of unmixing results obtained by applying our trained model on fundamental pattern images.

$$MP = \frac{1}{K(\Delta)} = \frac{\sigma_{\min}[\text{COV}(\Delta)]}{\sigma_{\max}[\text{COV}(\Delta)]}$$

$\Delta$  is a matrix where every column represents the object frequency or fluorescence fraction of a fundamental pattern.  $K(\Delta)$  is the condition number of  $\Delta$ .  $\sigma$  is the eigenvalue of the covariance matrix (the singular value of  $\Delta$ ). The model power range is  $0 \leq MP \leq 1$ , and a larger value corresponds to a more powerful model.

**Data and code availability.** All code and software used for the work described here is available from <http://murphylab.web.cmu.edu/software>.

**ACKNOWLEDGMENTS.** This work was supported in part by National Science Foundation Grant EF-0331657 (to R.F.M.) and National Institutes of Health Grants R01 GM075205 (to R.F.M.) and U54 RR022241 (to Alan Waggoner).

- Davis JR, Kakar M, Lim CS (2007) Controlling protein compartmentalization to overcome disease. *Pharm Res* 24:17–27.
- Kaffman A, O'Shea EK (1999) Regulation of nuclear localization: a key to a door. *Annu Rev Cell Dev Biol* 15:291–339.
- Schüller C, Ruis H (2002) Regulated nuclear transport. *Results Probl Cell Differ* 35:169–189.
- Kau TR, Way JC, Silver PA (2004) Nuclear transport and cancer: from mechanism to intervention. *Nat Rev Cancer* 4:106–117.
- Costes SV, et al. (2004) Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys J* 86:3993–4003.
- Comeau JWD, Costantino S, Wiseman PW (2006) A guide to accurate fluorescence microscopy colocalization measurements. *Biophys J* 91:4611–4622.
- Schubert W, et al. (2006) Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nat Biotechnol* 24:1270–1278.
- Dunlay RT, Czekalski WJ, Collins MA (2007) Overview of informatics for high content screening. *Methods Mol Biol* 356:269–280.
- Boland MV, Markey MK, Murphy RF (1998) Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. *Cytometry* 33:366–375.
- Murphy RF, Boland MV, Velliste M (2000) Towards a systematics for protein subcellular location: quantitative description of protein localization patterns and automated analysis of fluorescence microscope images. *Proc Int Conf Intell Syst Mol Biol* 8:251–259.
- Boland MV, Murphy RF (2001) A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics* 17:1213–1223.
- Conrad C, et al. (2004) Automatic identification of subcellular phenotypes on human cell arrays. *Genome Res* 14:1130–1136.
- Glory E, Murphy RF (2007) Automated subcellular location determination and high-throughput microscopy. *Dev Cell* 12:7–16.
- Hamilton NAPR, Pantelic RS, Hanson K, Teasdale RD (2007) Fast automated cell phenotype image classification. *BMC Bioinformatics* 8:110–117.
- Chen X, Velliste M, Murphy RF (2006) Automated interpretation of subcellular patterns in fluorescence microscope images for location proteomics. *Cytometry A* 69A:631–640.
- Zhao T, Velliste M, Boland MV, Murphy RF (2005) Object type recognition for automated analysis of protein subcellular location. *IEEE Trans Image Proc* 14:1351–1359.
- Bowman EJ, Siebers A, Altendorf K (1988) Bafilomycins: a class of inhibitors of membrane ATPases from microorganisms, animal cells, and plant cells. *Proc Natl Acad Sci USA* 85:7972–7976.
- Dröse S, et al. (1993) Inhibitory effect of modified bafilomycins and concanamycins on P- and V-type adenosinetriphosphatases. *Biochemistry* 32:3902–3906.
- Yamamoto A, et al. (1998) Bafilomycin A1 prevents maturation of autophagic vacuoles by inhibiting fusion between autophagosomes and lysosomes in rat hepatoma cell line, H-4-II-E cells. *Cell Struct Funct* 23:33–42.
- Ridler TW, Calvard S (1978) Picture thresholding using an iterative selection method. *IEEE Trans Syst Man Cybern* SMC-8:630–632.