

## Systems biology

# A hybrid machine-learning approach for segmentation of protein localization data

Peter M. Kasson<sup>1,2</sup>, Johannes B. Huppa<sup>5,6</sup>, Mark M. Davis<sup>5,6</sup> and Axel T. Brunger<sup>3,4,5,\*</sup>

<sup>1</sup>Biophysics Program, <sup>2</sup>Medical Scientist Training Program, <sup>3</sup>Department of Molecular and Cellular Physiology, Stanford Synchrotron Radiation Laboratory and <sup>4</sup>Department of Neurology and Neurological Sciences, Stanford Synchrotron Radiation Laboratory, Stanford University, Stanford, CA 94305, USA, <sup>5</sup>Howard Hughes Medical Institute, Stanford, CA 94305, USA and <sup>6</sup>Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA

Received on March 2, 2005; revised on May 30, 2005; accepted on August 4, 2005

Advance Access publication August 9, 2005

**ABSTRACT**

**Motivation:** Subcellular protein localization data are critical to the quantitative understanding of cellular function and regulation. Such data are acquired via observation and quantitative analysis of fluorescently labeled proteins in living cells. Differentiation of labeled protein from cellular artifacts remains an obstacle to accurate quantification. We have developed a novel hybrid machine-learning-based method to differentiate signal from artifact in membrane protein localization data by deriving positional information via surface fitting and combining this with fluorescence-intensity-based data to generate input for a support vector machine.

**Results:** We have employed this classifier to analyze signaling protein localization in T-cell activation. Our classifier displayed increased performance over previously available techniques, exhibiting both flexibility and adaptability: training on heterogeneous data yielded a general classifier with good overall performance; training on more specific data yielded an extremely high-performance specific classifier. We also demonstrate accurate automated learning utilizing additional experimental data.

**Availability:** <http://atb.slac.stanford.edu/~kasson/membraneclassifier.html>

**Contact:** [brunger@stanford.edu](mailto:brunger@stanford.edu)

**Supplementary information:** [http://atb.slac.stanford.edu/~kasson/classifier\\_suppl.pdf](http://atb.slac.stanford.edu/~kasson/classifier_suppl.pdf)

**INTRODUCTION**

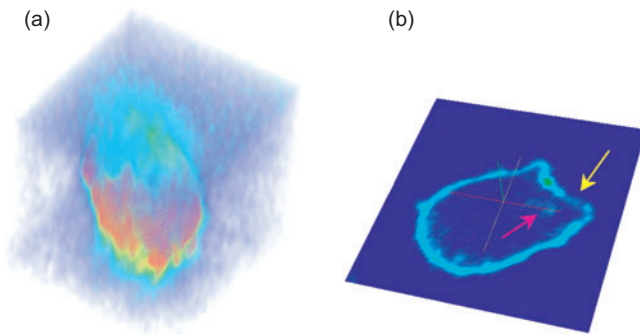
The use of subcellular protein localization data to provide positional information in addition to the expression data provided by genomic and proteomic assays promises to open new frontiers in the quantitative understanding of cellular function. Both static and dynamic localization of cell surface proteins control numerous aspects of cellular polarity, function, and signaling and defects in localization have been linked to the pathogenesis of neurodegenerative diseases

(Davies *et al.*, 1997; Saudou *et al.*, 1998), neuromuscular diseases (Gautam *et al.*, 1996; Ohno *et al.*, 2002), polycystic diseases (Huan and van Adelsberg, 1999; Roitbak *et al.*, 2004) and cancer metastasis (Singh *et al.*, 1998; Pagliarini and Xu, 2003; Xia *et al.*, 2004) among others. Protein localization is commonly monitored in real-time using fluorescence microscopy techniques. Using these microscopy images, quantitative information on protein localization changes can be extracted and combined with other functional data for quantitative analyses of the role of protein localization and redistribution in cellular function.

We and others have created systems to perform such analyses on specific classes of proteins (Gerlich *et al.*, 2001, 2003; Genovesio *et al.*, 2003; Kasson *et al.*, 2005), but the initial process of extracting the region of analytical interest (such as the cell membrane in the case of membrane protein localization) from fluorescence microscopy datasets remains a major obstacle to accurate analysis. This process, known as segmentation, is a critical stage because fluorescence signal that is erroneously excluded from the segmented region is lost to further analysis; signal that is erroneously included becomes noise. Further, erroneously included artifacts can change the topology of the extracted region, complicating subsequent spatial analyses. Currently available techniques suffer from both these weaknesses, and biological samples segmented with these techniques frequently require extensive manual correction before the results are suitable for accurate analysis. In this report, we present a novel, machine-learning-based hybrid method for segmentation of membrane protein localization data.

The problem of image segmentation—differentiating a structure of interest from the rest of the image—is often addressed by methods based on edge-finding region-filling or thresholding approaches. Segmentation of the plasma membrane from fluorescence microscopy images, however, is complicated by cellular autofluorescence and by the presence of internal accumulations of fluorophore (cytoplasmic inclusions, internalized dye or green fluorescent protein (GFP) fusion proteins present in the Golgi) as well as the problem of distinguishing membrane voxels from voxels external to

\*To whom correspondence should be addressed.



**Fig. 1.** Cell images from fluorescence microscopy of membrane proteins. Shown are visualizations of a CD3 $\zeta$ -GFP-labeled T cell. Displayed in (a) is a volume rendering; displayed in (b) is a midplane through the cell. Coloring is by voxel intensity. The magenta arrow in (b) indicates a region of intracellular fluorophore accumulation; the yellow arrow indicates a portion of the cell membrane that has low concentration of fluorophore.

the cell. A further challenge to segmentation is the fact that cell membranes in fluorescence microscopy images are often more than one voxel thick and may contain involutions (Fig. 1). Since the actual thickness of the cell membrane is substantially less than one voxel, this increased edge thickness may result from time-averaging of small positional fluctuations or may reflect increased local membrane curvature (Glebov and Nichols, 2004).

This combination of internal fluorophore accumulations and thick membrane edges results in inadequate performance by many traditional segmentation methods. To address these challenges, we have designed an advanced segmentation filter to be sensitive, specific and trainable to different types of data. Our hybrid classification algorithm extracts positional information regarding cell membrane contours using a level-set surface fitting approach (Caselles *et al.*, 1997) and then uses this information in conjunction with the original image to calculate a multidimensional set of feature vectors for each image voxel in the local neighborhood of the cell surface contour. Final classification is performed using a support vector machine (SVM) (Cortes and Vapnik, 1995; Joachims, 1999), a supervised learning technique for pattern recognition that has been successfully applied to fields such as text classification, object recognition (Pontil and Verri, 1998), medical diagnosis, and protein structure and function prediction (Bock and Gough, 2001; Ding and Dubchak, 2001; Chou and Cai, 2002; Karchin *et al.*, 2002).

We trained and evaluated our classifier on localization data for membrane signaling proteins involved in T lymphocyte activation. These proteins have been observed to undergo changes in localization specifically upon cellular activation (Monks *et al.*, 1998; Wulffing *et al.*, 1998; Grakoui *et al.*, 1999; Krummel *et al.*, 2000), and these changes are thought to have important functional consequences (Lee *et al.*, 2002; Davis *et al.*, 2003; Huppa and Davis, 2003; Huppa *et al.*, 2003; Lee *et al.*, 2003). Efficient segmentation of these localization data is required for accurate quantitative analysis of protein localization in a functional context. We trained and evaluated this classifier on both heterogeneous and more narrowly selected image data, showing substantial gains over existing segmentation methods for each. We have also demonstrated the ability of our classifier to learn based on dual-color data that use a membrane-specific probe to incorporate an experimental standard for membrane localization.

## SYSTEM AND METHODS

### Cell preparation and image acquisition

5C.C7 T lymphoblasts derived from T-cell-receptor transgenic mouse lymph nodes were grown, retrovirally transfected with fluorophore fusion protein constructs, and stimulated *in vitro* with the moth cytochrome *c* peptide 88–103 presented on antigen-presenting cells (CH27/I-E<sup>k</sup>) as reported previously (Fink *et al.*, 1986; Ehrlich *et al.*, 2002; Huppa *et al.*, 2003). Expert manual segmentation experiments (see below) were performed using cells transfected with CD3 $\zeta$ -GFP or LAT-GFP fusion proteins, while membrane probe dual-color segmentation experiments were performed using cells co-transfected with CD3 $\zeta$ -GFP and PKC $\delta$ -PH-YFP fusion proteins. Images were acquired on a Zeiss Axiovert S100TV microscope with a 1.3 NA 40 $\times$  Neo-Fluar or Fluar objective or a 1.4 NA 63 $\times$  Apochromat objective (Carl Zeiss, Jena, Germany). Samples were illuminated by a 300 W xenon light source with a Sutter DG-4 filter changer (Sutter Instruments, Novato, CA). Detection was performed using a cooled charge coupled device (CCD) camera (Roper Scientific, Tucson, AZ). Z-scanning was accomplished using a piezo-driven motor (Physik Instrumente, Waldbronn, Germany). Cells were imaged at 37 $^{\circ}$ C in phenol red-free RPMI. Metamorph 5.0 (Universal Imaging Corporation, Downingtown, PA) was used for microscope control; images were further processed using blind deconvolution (Deblur 9.2; AutoQuant Imaging, Watervliet, NY). Datasets were collected at a resolution of 0.3  $\mu$ m  $\times$  0.3  $\mu$ m  $\times$  1  $\mu$ m for dual-color experiments and 0.5  $\mu$ m  $\times$  0.5  $\mu$ m  $\times$  1  $\mu$ m for single-color experiments. These resolutions were chosen to optimize the trade-off between high-resolution data and the motion blurring that results from slower, higher-resolution imaging of live cells. Image datasets are enumerated in Supplementary Table 1.

### Initial segmentation and calculation of SVM inputs

Our segmentation algorithm is schematized in Figure 2. The learning-based classifier utilizes a surface-fitting method to obtain an initial approximate segmentation of the cell surface (Fig. 2a). For this purpose we employ the active contour (level-set-based) model of Caselles *et al.* (1997). Voxels within the local neighborhood of the surface, empirically set as 1.7  $\mu$ m, are designated for inclusion in the initial segmentation (Fig. 2b). Inputs to the SVM module are then calculated for these voxels (Fig. 2c). Euclidean distance from the surface is determined by calculation of a distance map from the active contour surface as per Danielsson (1980). The observed voxel intensity and voxel intensity gradient magnitude are calculated directly from the deconvolved microscopy data. Finally, the surface normal vector is determined for each voxel in the initial segmentation and the intensity gradient magnitude is calculated for all voxels in the local neighborhood along that vector. These four inputs constitute the input vector space for the SVM module (Fig. 2d).

### SVM module

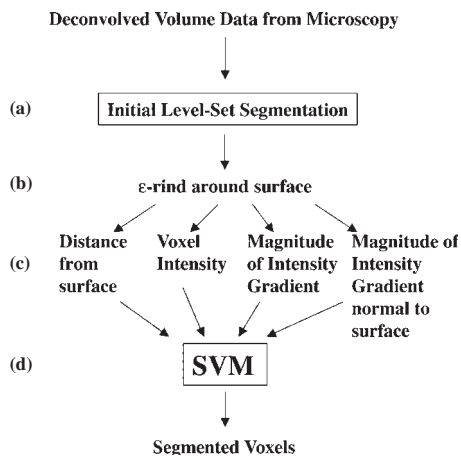
We employ an SVM (using the SVMlight implementation) using a radial basis kernel function as follows:

$$f(x) = \sum_j G\left(\frac{\|x - \xi_j\|}{\lambda_j}\right) \beta_j, \quad (1)$$

where  $f(x)$  is the function in transformed space,  $G$  is a Gaussian density function,  $x$  is the input vector,  $\xi_j$  are the support vectors,  $\lambda_j$  are the scaling parameters,  $\beta_j$  are the fit parameters (Cortes and Vapnik, 1995; Joachims, 1999). Both first- and second-order kernels were tested, letting the kernel function be either a polynomial or a radial basis function. A first-order radial basis function was selected based on optimal performance on the training set. The learning parameter  $c$  was set to 0.1.

### SVM training

We have pursued two approaches for training the learning-based membrane classifier. The first and more flexible is training via manual segmentation by experts. To obtain training sets via expert manual segmentation, biologists



**Fig. 2.** Learning-based approach for volume segmentation. This schema gives a high-level depiction of the supervised learning algorithm for volume segmentation. In stage (a), a level-set surface-fitting algorithm is used to produce an initial estimate of the cell surface from the deconvolved microscopy data. The algorithm employed is a geodesic active contour segmentation approach (Caselles *et al.*, 1997). In (b), an  $\varepsilon$ -rind is taken around the surface (using  $\varepsilon = 5$  voxels). In (c), the voxels selected in (b) are used to compute the following inputs for the SVM module: distance from the initial surface, voxel intensity, magnitude of the intensity gradient and magnitude of the intensity gradient normal to the initial surface. In (d), the SVM module [(SVM implementation by Joachims (1999))] classifies each voxel selected in (b) as either membrane or non-membrane, yielding a set of segmented voxels. A radial basis function kernel is used for the SVM, letting the function order  $\gamma = 1$ ,  $C = 0.1$ .

who regularly analyze T-cell images were asked to trace cell contours for a series of cells, each represented as a series of planar slices in two different orientations (sequential  $x$ - $y$  and  $x$ - $z$  planes, respectively). Each slice was traced four times: by two independent biologists in two orientations each. These tracings were scanned and combined to yield a segmentation confidence score for each voxel. This score was converted to a binary classification by a majority vote scheme, and the resulting segmentation was used to train the SVM.

The five images in the heterogeneous training set were selected for segmentation and training based on their diversity: these images were acquired on four different days and represent cells labeled with two different fluorescent probe constructs: CD3 $\zeta$ -GFP and LAT-GFP. To create a more specific training set, images were selected from experiments on cells expressing LAT-GFP. Only images collected on the same day and manually segmented on the same day were used; three cells fit these criteria.

The second and more powerful approach for training our membrane classifier is to perform two-color microscopy experiments using a fluorescent probe for the protein of interest in conjunction with a fluorescent membrane probe. We have performed these experiments using a cyan fluorescent protein (CFP)-CD3 $\zeta$  conjugate that co-migrates with the T-cell receptor and a yellow fluorescent protein (YFP)-protein kinase C $\delta$  pleckstrin homology domain (PKC $\delta$ -PH) conjugate that uniformly labels the plasma membrane (Stauffer *et al.*, 1998; Varnai and Balla, 1998). Training based on dual-color fluorescence with a membrane probe was performed using the membrane probe to derive a cell-membrane binary classification for use in SVM training on the test probe (in this case CD3 $\zeta$ ) data. The membrane probe data were classified via  $k$ -means segmentation of the deconvolved microscopy data ( $k = 3$ ; the highest-intensity group is selected) or via segmentation of the deconvolved microscopy data using the Moss filter. This classification was used to train the SVM.

## Implementation of previously available classifiers

$k$ -means clustering was performed using the Matlab Statistics Toolbox (The MathWorks, Natick, MA). Canny filtering (Canny, 1986) and level-set surface fitting (Caselles *et al.*, 1997) were implemented using the ITK class library (<http://www.itk.org>). The Moss filter (Yang *et al.*, 2001; Moss *et al.*, 2002) was implemented based on code provided by William Moss.

## Evaluation of performance

Performance statistics are computed as follows, using expert manual segmentation as a reference standard. Accuracy is defined as

$$\left( \frac{TP + TN}{TP + FP + TN + FN} \right),$$

precision as

$$\left( \frac{TP}{TP + FP} \right)$$

and recall as

$$\left( \frac{TP}{TP + FN} \right),$$

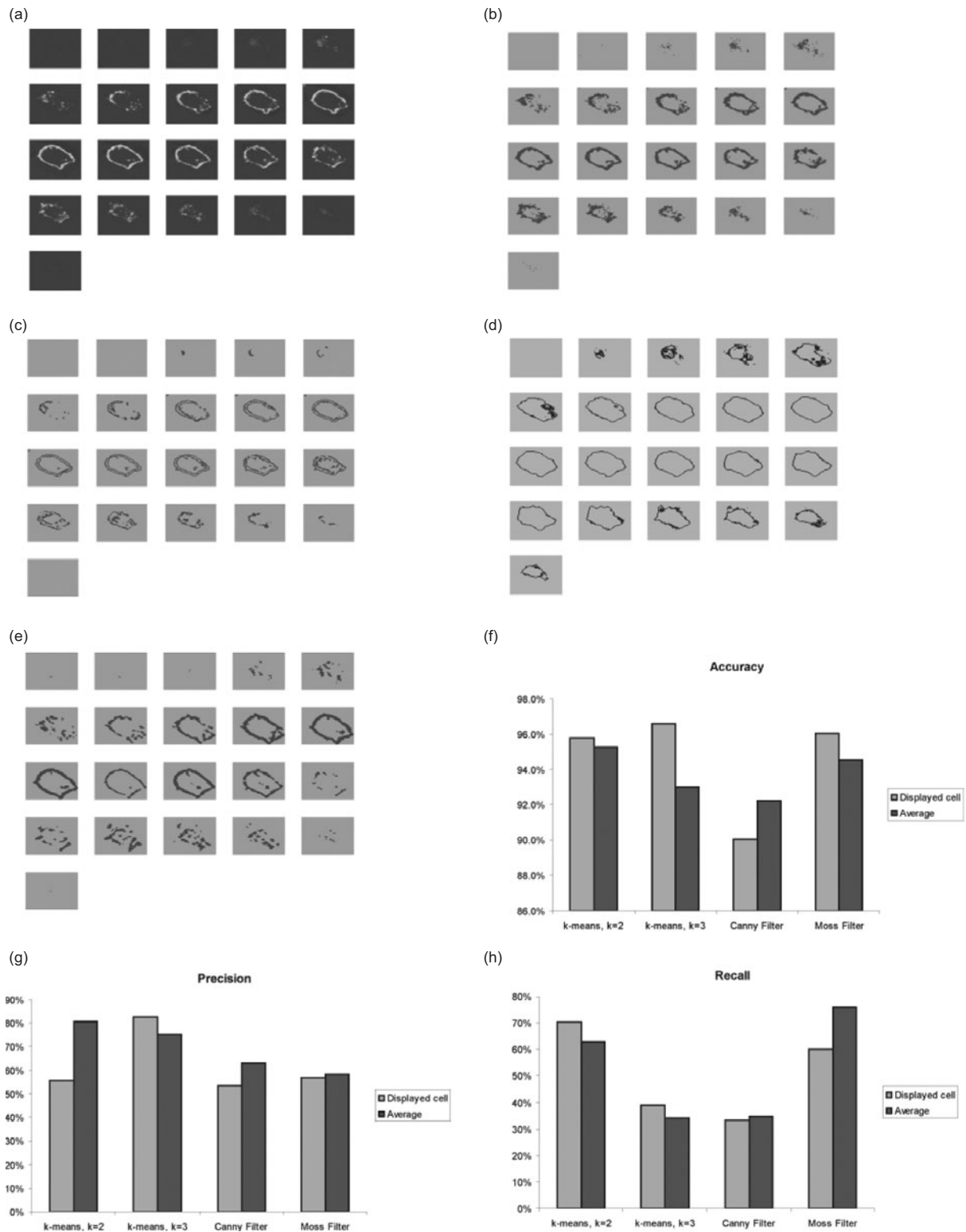
where TP denotes the number of true positive classifications, FP the number of false positives, TN the number of true negatives and FN the number of false negatives. Accuracy scores provide a composite measure of successful membrane classification and successful exclusion of non-membrane voxels. Recall scores measure how well membrane voxels were detected by the classifier, and precision scores provide confidence values for positive results. For measurements of intracellular artifact inclusion, distances to the center of the cell were computed for all voxels. Voxels  $x$  for which  $d_c(\text{surfnearest}(x)) - d_c(x) > \text{one voxel width}$  (0.3 or 0.5  $\mu\text{m}$ ) were specified as intracellular, where  $d_c$  is the distance to the center of the cell and  $\text{surfnearest}(x)$  is the surface voxel nearest to  $x$ .

## RESULTS AND DISCUSSION

We have used localization data from membrane signaling proteins involved in T-cell activation to test and evaluate our learning-based classifier. We first compare the performance of our classifier to that of previously available methods on a heterogeneous set of experimental data and then go on to perform similar comparisons for more narrowly focused data. In both cases, the learning-based classifier out-performs other available methods, particularly with respect to the elimination of intracellular artifacts. We also demonstrate methods for training the learning-based classifier using additional fluorescent labels to visualize the cell membrane. These multiple training regimes illustrate the power and flexibility of our approach, as the learning-based classifier can easily be trained for optimal performance in a wide variety of applications.

### Comparison of previously existing classification methods

Representative segmentations of a membrane protein localization dataset using a number of segmentation methods and comparisons of their performance statistics (using expert manual segmentation as a reference) are displayed in Figure 3. Adaptive thresholding methods such as  $k$ -means clustering perform reasonably well on cell membranes (Fig. 3b) but, because they ignore positional information, they erroneously include any high-intensity artifacts present in the image. Many edge-based methods such as Canny filtering perform less well on fluorescence microscopy images because of the thick membrane edges (Fig. 3c). The Canny filter produces a relatively poor segmentation, particularly in terms of accuracy and recall—it does not capture membrane voxels well. More robust surface fitting



**Fig. 3.** Existing methods for cell membrane segmentation. Shown in (a) is the cell displayed in Figure 1, rendered as a series of sequential  $x$ - $y$  planes through the volume dataset. Shown in (b) is the inclusion mask for the same cell segmented via  $k$ -means ( $k = 2$ ) clustering on intensity. Shown in (c) is a similar inclusion mask for segmentation by Canny filtering, with the threshold value for thinning set equal to the image median. This value was determined to have near-optimal accuracy on a single test image. Shown in (d) the output of level-set surface fitting, and shown in (e) is the inclusion mask for segmentation via the Moss filter. Plotted in (f) is the accuracy of the segmentation filters, plotted in (g) is the precision, and plotted in (h) is the recall with respect to manual segmentation.

approaches have been used successfully for related problems in biological image segmentation (Zimmer *et al.*, 2002). Such approaches can successfully segment cell contours, but they fail to capture membrane thickness and involutions (Fig. 3d). The Moss filter (Yang *et al.*, 2001; Moss *et al.*, 2002), which was developed specifically for the segmentation of membrane structures, detects membrane structures with fairly good sensitivity, successfully capturing thick membrane edges (Fig. 3e). It nevertheless has two major drawbacks. First, it produces a segmented cell membrane that is not fully connected and often has substantial holes. Second, it often includes intracellular artifacts that can substantially distort analysis. The Moss and *k*-means filters have approximately equivalent overall performance. Intracellular artifacts are also included to a similar extent by the *k*-means and Moss filters. Common challenges to all of the methods compared include overall performance, membrane thickness and particularly the erroneous inclusion of intracellular artifacts.

### General classifier

Visualizations from representative output of the learning-based classifier trained on expert manual segmentations are shown in Figures 4a and b. Accuracy, precision and recall for the learning-based classifier were determined via 5-fold cross-validation on the training data and compared with the performance of the Moss filter and *k*-means clustering on the same data. As can be seen from the figures, performance of the learning-based classifier (and that of all other classifiers tested) is best in the planes closest to the equator of the cell, where the fluorescence signal is highest. Intracellular artifacts are virtually absent from the segmentation generated by the learning-based classifier.

A comparison of the performance of the expert-manual-segmentation-trained classifier with that of the Moss filter is given in Figure 4c–e. As can be seen from the performance statistics, the learning-based classifier has improved accuracy over both the Moss filter and *k*-means clustering on the test data, slightly increased precision and comparable recall. Performance variations between datasets result from several factors. First, a relatively heterogeneous set of test data were chosen. Datasets from both CD3 $\zeta$ -GFP-labeled and LAT-GFP-labeled cells were used; the fluorophore signal quality also varied substantially between datasets. This heterogeneous set was chosen to train a classifier that would be as general as possible; a classifier trained on a more narrowly chosen training set, as shown in the next section, will have better performance on images similar to that training set. Also, the quality of the expert segmentations varies somewhat from test image to test image. This is linked to a larger issue—training and validation by expert segmentation is only as good as the expert segmentations themselves, and not all imaging experts would create the same segmentation given the same image. Nonetheless, the ability of the learning-based classifier to be trained by biologists themselves to suit their particular purposes provides flexibility and applicability to a wide range of biological imaging problems.

One of the motivations for designing a novel segmentation filter was to reduce the number of intracellular artifacts segmented as membrane. Figure 4f shows a comparison of the learning-based classifier with the Moss filter and *k*-means classification in this respect. The learning-based filter labels ~10-fold fewer intracellular points as membrane than either the Moss filter or *k*-means segmentation. Similarly, the false positive points (those labeled membrane by the classifier but not in the expert segmentation) identified by the

learning-based classifier were on average 64% closer to the nearest expert-labeled membrane point (and thus to the cell surface) than those identified by the Moss filter and 210% closer than those identified by *k*-means clustering (Supplemental Figure 1b). Most false positives identified by either the learning-based classifier or the Moss filter were in close proximity to the cell membrane, but this situation occurs to a greater extent in the learning-based classified output. The *k*-means clustering output was notable for having a substantial number of false positive points far from the membrane. A visualization of the expert-segmented membrane points and the intracellular points labeled membrane by each of the learning-based classifier and Moss filter is shown in Supplemental Figure 1c.

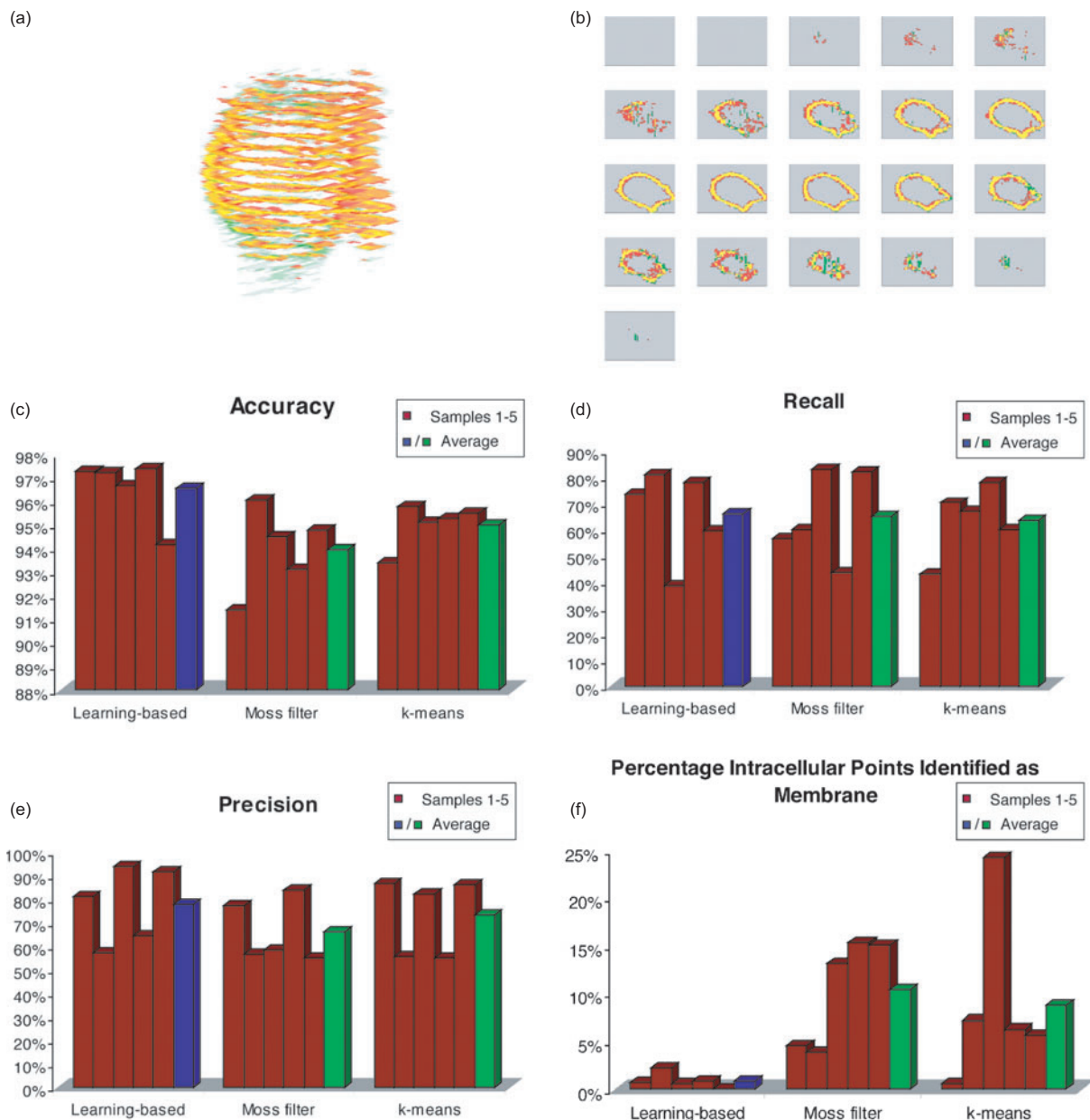
One weakness of all the classifiers tested, including both pre-existing segmentation methods and the learning-based classifier described in this work, is that the ‘top’ and ‘bottom’ of the cell as it sits on the microscopy stage are segmented substantially less well than planes near the equator. This difference results from lower fluorescence signal at the poles of the cell, which in turn is caused in part by anisotropy of the image voxels (asymmetry resulting from image resolution differing in the *z*-dimensions from that in *x* and *y*). Fluorescence signal is increased near the equator of the cell where the membrane tangent plane is parallel to the long axis of each voxel, and is decreased near the poles where the membrane tangent plane is orthogonal to the long axis of each voxel. Ideally, a classifier would take into account the direction of the membrane tangent relative to the unit voxel dimensions in order to correct for this problem. This remains an area for future development and a problem for which machine-learning approaches are well suited.

### LAT-specific classifier

Owing to the nature of supervised learning, it is expected that training and cross-validation on a more narrowly chosen range of images will yield a greater gain in performance. To demonstrate this more selective training scheme, images of three cells were chosen from a single dataset of LAT-GFP-labeled T cells stimulated as described in the System and Methods section. The results of cross-validation testing on these cells are shown in Figure 5. As can be seen in the figure, the LAT-specific classifier has a 35% point improvement in precision, a 6% point improvement in recall, and a 2% point improvement in accuracy compared with the Moss filter. Compared with *k*-means clustering, it has a 14% point improvement in precision, a 17% point improvement in recall, and a 2% point improvement in accuracy. Our learning-based approach is thus quite flexible; it can be used to train a generally applicable classifier with a moderate gain in performance over existing segmentation algorithms, or it can be used to train a more situation-specific classifier with markedly increased performance.

### Automated training from membrane probe data

Even in the case of expert manual segmentation of the plasma membrane, it is often challenging to differentiate involutions in the plasma membrane from intracellular inclusions. Expert manual segmentation is also time-consuming and somewhat variable from inspection to inspection and expert to expert. The ability to train a classifier based on an experimental standard for the plasma membrane holds the potential to address these drawbacks to manual segmentation. To investigate this possibility, a number of imaging experiments were

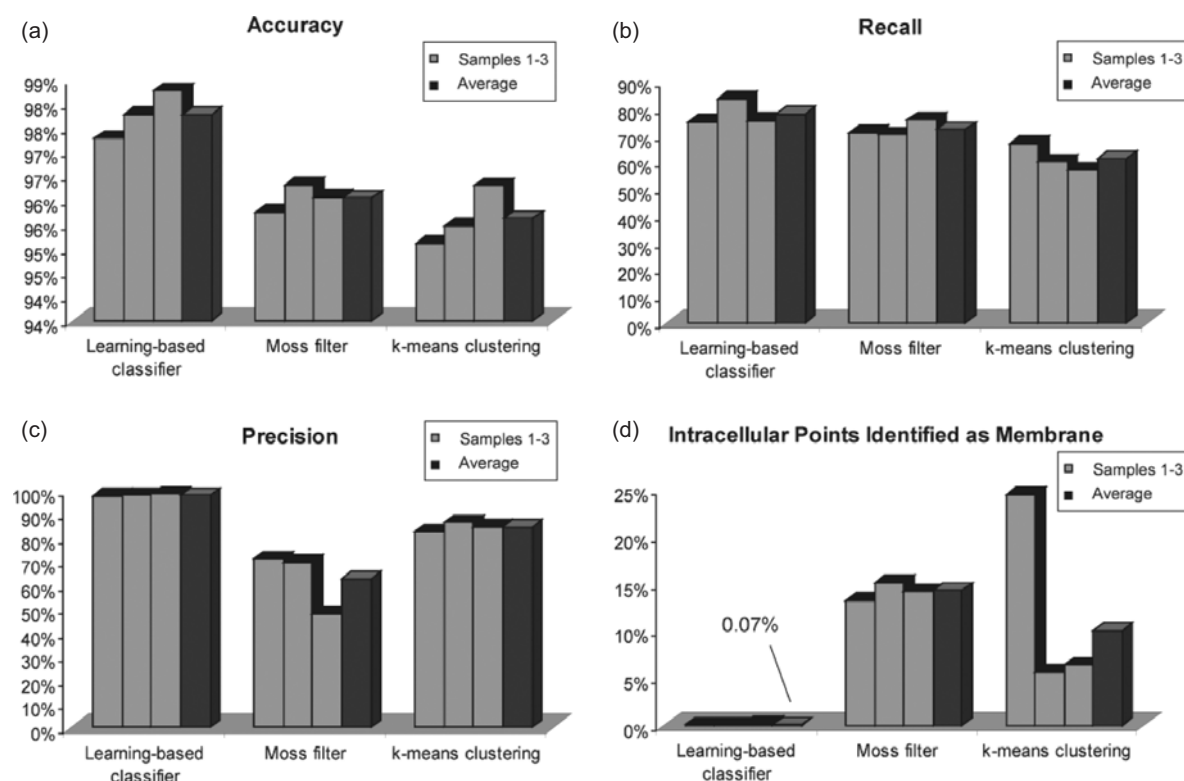


**Fig. 4.** Learning-based segmentation of the plasma membrane. Shown are the results of 5-fold cross-validation experiments for the training of our membrane classifier. Using a set of five 3D microscopy images of cells, the classifier was trained on four of the five and tested on the fifth. This procedure was repeated for each cell in the test set. The test set contained three cells expressing CD3 $\zeta$ -GFP and two cells expressing LAT-GFP. Rendered in (a) is a comparison of our learning-based segmentation (in red) with manual segmentation (in green) of GFP-labeled LAT on the plasma membrane of a single T lymphocyte. Regions of overlap are in yellow. The rendering depicts sequential 1  $\mu$ m slices through the cell volume. The same slices are given as sequential  $x$ - $y$  planes in (b). (c-e) The performance of our learning-based classifier with that of the Moss segmentation filter and  $k$ -means clustering with respect to accuracy of detection (c), precision (d) and recall (e) are compared. Plotted in (f) is the percentage of intracellular points erroneously identified as membrane (with respect to manual segmentation) by each classification method. For (c-f), both the individual cross-validation results and the average results are plotted.

performed in which T cells were transfected with both a CD3 $\zeta$ -CFP probe and a membrane-specific PLC $\delta$ -PH-YFP probe.

As a first step in the analysis of the dual-fluorophore data, the membrane probe data were automatically segmented to yield a reference set for training the learning-based classifier on the CD3 $\zeta$  data. Two approaches were pursued for this automatic segmentation:  $k$ -means

clustering ( $k = 3$ , with the highest-intensity cluster labeled as membrane) and Moss filtering. The learning-based classifier was trained with each approach and the results compared with those from the learning-based classifier trained on a manual segmentation of the membrane probe data and those from the Moss filter and  $k$ -means clustering applied to the CD3 $\zeta$  data. Performance statistics



**Fig. 5.** Performance of a LAT-specific learning-based classifier. A learning-based classifier trained on a dataset of LAT-GFP-labeled T cell images performs substantially better on images of other LAT-GFP-labeled T cells. Shown in this figure are the results of cross-validation testing on a LAT dataset. Plotted in (a) is the accuracy of the LAT-specific learning-based classifier compared with that of the Moss filter and *k*-means filtering ( $k = 2$ ), plotted in (b) is the precision, and plotted in (c) is the recall. The percentage of erroneously included intracellular voxels is plotted in (d). For each panel, both the individual cross-validation results and the average results are plotted.

and visualizations comparing these segmentation methods are given in Figure 6. The *k*-means-trained classifier and the manually trained classifier had the best overall performance. Interestingly, the Moss-trained classifier is similar to the Moss filter in its performance characteristics, but the *k*-means-trained classifier improves performance over *k*-means segmentation directly on the CD3 $\zeta$  data. Training this latter classifier using the high-precision 3-means classification most probably provides some advantage. The *k*-means-trained classifier is thus able to score within measurement error of the manually trained classifier.

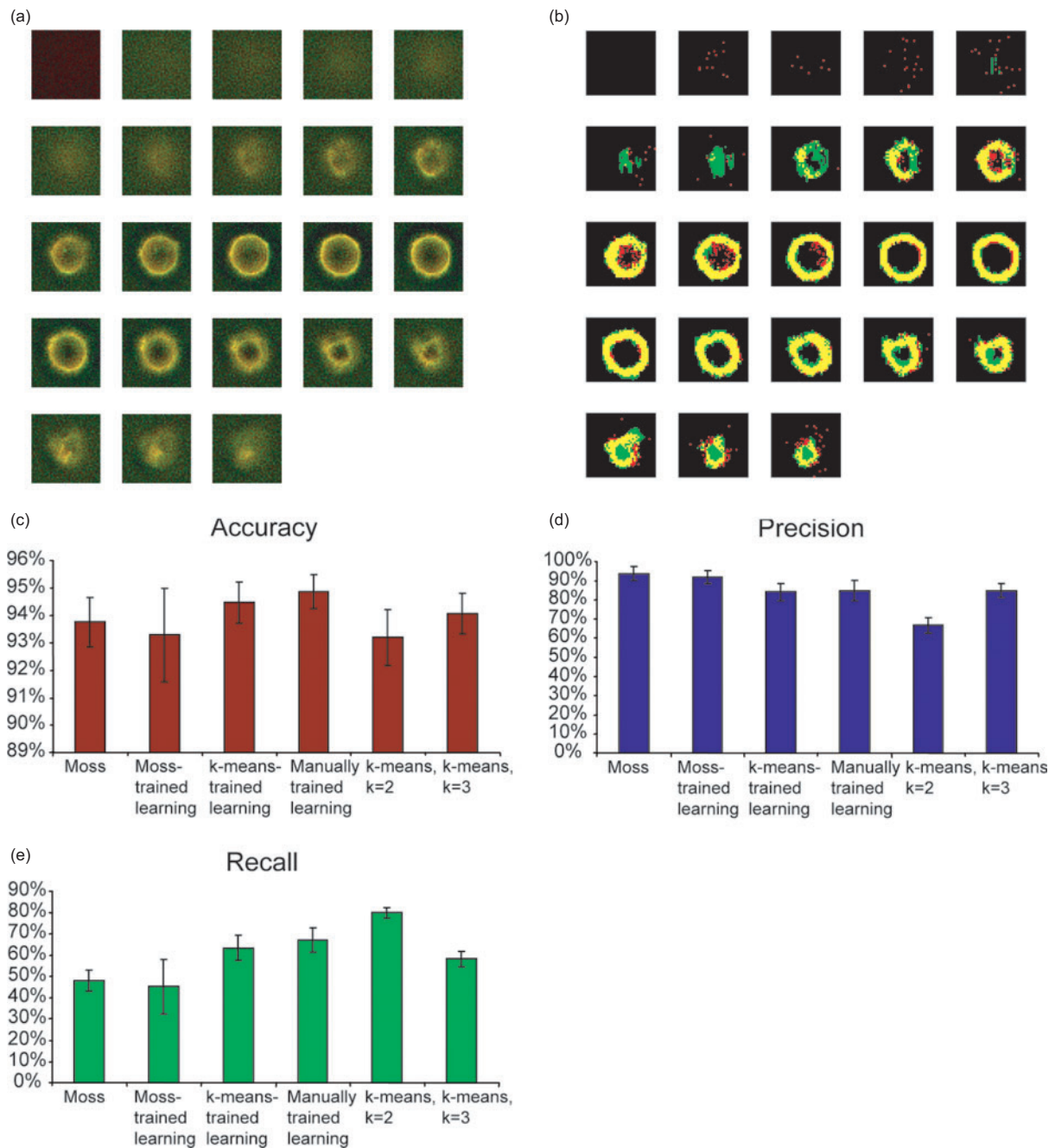
Part of the difficulty in obtaining results from automatically trained classifiers equivalent to those from a manually trained classifier stems from the inherent circularity of the automatic training problem. One has to segment the membrane probe data in order to train the classifier to segment the data for the labeled protein of interest. Fortunately, segmentation of the membrane probe data is a substantially easier problem. The membrane probe used in these studies, PKC $\delta$ -PH-YFP, localizes very specifically to the plasma membrane, without the intracellular inclusions observed with proteins such as CD3 $\zeta$  and LAT. The fact that the *k*-means-trained classifier performance is increased over the performance of direct *k*-means clustering on the CD3 $\zeta$  data is indicative of this reduction in scope of the classification problem.

Beyond automated training of the classifier, an additional benefit of working with dual-fluorophore data with a membrane probe is

the ability better to differentiate membrane and membrane-proximal labeled protein from internalized labeled protein. Such internalization is particularly prominent and important in T cells undergoing activation. As currently implemented, our methods can only make this distinction accurately to within the resolution of the image data; however, use of additional imaging techniques such as fluorescence resonance energy transfer could improve the resolution of differentiation by an order of magnitude. In addition, since our dual-fluorophore data come from resting cells, clarification of the nature of membrane involutions in activated T cells remains an area for future research.

## CONCLUSIONS

We have developed a novel learning-based method for the classification of plasma membrane protein localization data obtained via fluorescence microscopy and the differentiation of these data from intracellular artifacts. The results we have presented demonstrate the increased performance of a learning-based classifier over existing methods for membrane segmentation. The primary advantages of this classifier are its flexibility and adaptability. Training the classifier on data from experimental conditions similar to the test data yields an extremely good classifier for those test data. Because it is often not desirable to re-train the classifier for every type of experimental data, we have also trained our classifier on heterogeneous



**Fig. 6.** Segmentation of dual-color fluorescence images. Dual-color fluorescence images were segmented using the CD3 $\zeta$ -CFP test probe channel as input and the PLC $\delta$ -PH-YFP membrane probe channel for training and reference segmentations. Displayed in (a) is an overlay of the membrane probe microscopy data in red and the test probe microscopy data in green. Displayed in (b) is an overlay of the reference segmentation in green and the output of the learning-based classifier automatically trained on the data using  $k$ -means segmentation in red. Multiple images represent sequential  $x$ - $y$  planes through the volume data. Plotted in (c) is a comparison of the accuracy of different segmentation methods for the test probe data. The precision is similarly plotted in (d), and the recall is plotted in (e). Values shown are averages over the test set ( $n = 5$ ), and error bars represent 1 SD of the mean.

data and demonstrated performance on such data superior to that of other available methods. Even a classifier trained to be extremely general in its recognition abilities showed performance gains in several key areas, particularly the removal of intracellular artifacts and overall improvements in accuracy. Much of this inherent benefit

to the learning-based classifier derives from its integration of several types of image data: intensity information, surface positional information and gradient edge information. Because analyses such as those of protein clustering that we have performed in earlier studies require membranes to be segmented with minimal artifacts, our new



learning-based classifier increases both the accuracy of the analyses and makes them more automated. We have applied this classifier to the study of protein localization during T-cell activation and combined the resulting protein localization data with functional data to study T-cell signaling in an approach that generalizes to a range of signaling phenomena.

## ACKNOWLEDGEMENTS

We thank O. Troyanskaya, M. Vrljic and T. Fenn for many helpful discussions. W. Moss allowed use of code for the Moss Filter. This work was supported in part by the National Institutes of Health (M.M.D.) and the Medical Scientist Training Program. Funding to pay the Open Access publication charges for this article was provided by the Howard Hughes Medical Institute.

*Conflict of Interest:* none declared.

## REFERENCES

- Bock, J.R. and Gough, D.A. (2001) Predicting protein-protein interactions from primary structure. *Bioinformatics*, **17**, 455–460.
- Canny, J. (1986) A computational approach to edge-detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, **8**, 679–698.
- Caselles, V. et al. (1997) Geodesic active contours. *Int. J. Comput. Vision*, **22**, 61–79.
- Chou, K.C. and Cai, Y.D. (2002) Using functional domain composition and support vector machines for prediction of protein subcellular location. *J. Biol. Chem.*, **277**, 45765–45769.
- Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Mach. Learn.*, **20**, 273–297.
- Danielsson, P.E. (1980) Euclidean distance mapping. *Comput. Graph. Image Process.*, **14**, 227–248.
- Davies, S.W. et al. (1997) Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell*, **90**, 537–548.
- Davis, M.M. et al. (2003) Dynamics of cell surface molecules during T cell recognition. *Annu. Rev. Biochem.*, **72**, 717–742.
- Ding, C.H.Q. and Dubchak, I. (2001) Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics*, **17**, 349–358.
- Ehrlich, L.I. et al. (2002) Dynamics of p56lck translocation to the T cell immunological synapse following agonist and antagonist stimulation. *Immunity*, **17**, 809–822.
- Fink, P.J. et al. (1986) Correlations between T-cell specificity and the structure of the antigen receptor. *Nature*, **321**, 219–226.
- Gautam, M. et al. (1996) Defective neuromuscular synaptogenesis in agrin-deficient mutant mice. *Cell*, **85**, 525–535.
- Genovesio, A., Zhang, B. et al. (2003) Tracking of multiple fluorescent biological objects in three dimensional video microscopy. In *2003 International Conference on Image Processing*.
- Gerlich, D. et al. (2001) Four-dimensional imaging and quantitative reconstruction to analyse complex spatiotemporal processes in live cells. *Nat. Cell Biol.*, **3**, 852–855.
- Gerlich, D. et al. (2003) Quantitative motion analysis and visualization of cellular structures. *Methods*, **29**, 3–13.
- Glebov, O.O. and Nichols, B.J. (2004) Lipid raft proteins have a random distribution during localized activation of the T-cell receptor. *Nat. Cell Biol.*, **6**, 238–243.
- Grakoui, A. et al. (1999) The immunological synapse: a molecular machine controlling T cell activation. *Science*, **285**, 221–227.
- Huan, Y. and van Adelsberg, J. (1999) Polycystin-1, the PKD1 gene product, is in a complex containing E-cadherin and the catenins. *J. Clin. Invest.*, **104**, 1459–1468.
- Huppa, J.B. and Davis, M.M. (2003) T-cell-antigen recognition and the immunological synapse. *Nat. Rev. Immunol.*, **3**, 973–983.
- Huppa, J.B. et al. (2003) Continuous T cell receptor signaling required for synapse maintenance and full effector potential. *Nat. Immunol.*, **4**, 749–755.
- Joachims, T. (1999) Making large-scale SVM learning practical. In Scholkopf, B., Burges, C. and Smola, A. (eds) *Advances in Kernel Methods—Support Vector Learning*. MIT Press, Boston.
- Karchin, R. et al. (2002) Classifying G-protein coupled receptors with support vector machines. *Bioinformatics*, **18**, 147–159.
- Kasson, P.M. et al. (2005) Quantitative imaging of lymphocyte membrane protein reorganization and signaling. *Biophys. J.*, **88**, 579–589.
- Krummel, M.F. et al. (2000) Differential clustering of CD4 and CD3zeta during T cell recognition. *Science*, **289**, 1349–1352.
- Lee, K.H. et al. (2002) T cell receptor signaling precedes immunological synapse formation. *Science*, **295**, 1539–1542.
- Lee, K.H. et al. (2003) The immunological synapse balances T cell receptor signaling and degradation. *Science*, **302**, 1218–1222.
- Monks, C.R. et al. (1998) Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature*, **395**, 82–86.
- Moss, W.C. et al. (2002) Quantifying signaling-induced reorientation of T cell receptors during immunological synapse formation. *Proc. Natl Acad. Sci. USA*, **99**, 15024–15029.
- Ohno, K. et al. (2002) Rapsyn mutations in humans cause endplate acetylcholine-receptor deficiency and myasthenic syndrome. *Am. J. Hum. Genet.*, **70**, 875–885.
- Pagliarini, R.A. and Xu, T. (2003) A genetic screen in *Drosophila* for metastatic behavior. *Science*, **302**, 1227–1231.
- Pontil, M. and Verri, A. (1998) Support vector machines for 3D object recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, **20**, 637–646.
- Roitbak, T. et al. (2004) A polycystin-1 multiprotein complex is disrupted in polycystic kidney disease cells. *Mol. Biol. Cell.*, **15**, 1334–1346.
- Saudou, F. et al. (1998) Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell*, **95**, 55–66.
- Singh, S.P. et al. (1998) Loss or altered subcellular localization of p27 in Barrett's associated adenocarcinoma. *Cancer Res.*, **58**, 1730–1735.
- Stauffer, T.P. et al. (1998) Receptor-induced transient reduction in plasma membrane PtdIns(4,5)P2 concentration monitored in living cells. *Curr. Biol.*, **8**, 343–346.
- Varnai, P. and Balla, T. (1998) Visualization of phosphoinositides that bind pleckstrin homology domains: calcium- and agonist-induced dynamic changes and relationship to myo-[3H]inositol-labeled phosphoinositide pools. *J. Cell Biol.*, **143**, 501–510.
- Wulfing, C. et al. (1998) Visualizing the dynamics of T cell activation: intracellular adhesion molecule 1 migrates rapidly to the T cell/B cell interface and acts to sustain calcium levels. *Proc. Natl Acad. Sci. USA*, **95**, 6302–6307.
- Xia, W. et al. (2004) Phosphorylation/cytoplasmic localization of p21Cip1/WAF1 is associated with HER2/neu overexpression and provides a novel combination predictor for poor prognosis in breast cancer patients. *Clin. Cancer Res.*, **10**, 3815–3824.
- Yang, J. et al. (2001) Telomerized human microvasculature is functional *in vivo*. *Nat. Biotechnol.*, **19**, 219–224.
- Zimmer, C. et al. (2002) Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: a tool for cell-based drug testing. *IEEE Trans. Med. Imaging*, **21**, 1212–1221.