A spectral graph theoretic approach to quantification and calibration of collective morphological differences in cell images

Yu-Shi Lin[1,2,†], Chung-Chih Lin[3,†], Yuh-Show Tsai[4,], Tien-Chuan Ku[4], Yi-Hung Huang[1,2] and Chun-Nan Hsu[1,5,*]

[1]Institute of Information Science, Academia Sinica, [2]Department of Computer Science and Information Eng., National Taiwan University, [3]Department of Life Science and Institute of Genome Science, National Yang-Ming University, Taipei, [4]Department of Biomedical Engineering, Chung Yuan Christian University, Jhongli, Taiwan and [5]Information Sciences Institute, University of Southern California, Marina del Rey, CA, USA

ABSTRACT

Motivation: High-throughput image-based assay technologies can rapidly produce a large number of cell images for drug screening, but data analysis is still a major bottleneck that limits their utility. Quantifying a wide variety of morphological differences observed in cell images under different drug influences is still a challenging task because the result can be highly sensitive to sampling and noise.

Results: We propose a graph-based approach to cell image analysis. We define graph transition energy to quantify morphological differences between image sets. A spectral graph theoretic regularization is applied to transform the feature space based on training examples of extremely different images to calibrate the quantification. Calibration is essential for a practical quantification method because we need to measure the confidence of the quantification. We applied our method to quantify the degree of partial fragmentation of mitochondria in collections of fluorescent cell images. We show that with transformation, the quantification can be more accurate and sensitive than that without transformation. We also show that our method outperforms competing methods, including neighbourhood component analysis and the multi-variate drug profiling method by Loo *et al.* We illustrate its utility with a study of Annonaceous acetogenins, a family of compounds with drug potential. Our result reveals that squamocin induces more fragmented mitochondria than muricin A.

Availability: Mitochondrial cell images, their corresponding feature sets (SSLF and WSLF) and the source code of our proposed method are available at http://aiia.iis.sinica.edu.tw/.

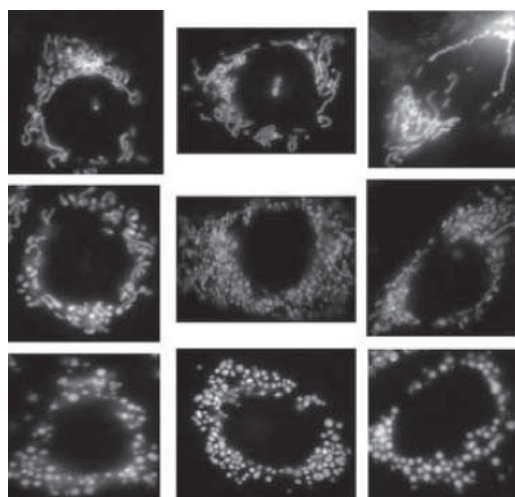**Contact:** chunnan@iis.sinica.edu.tw

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Recently, high-throughput image-based assay technologies, or high-content analysis, have become a useful tool for drug discovery (Jones *et al.*, 2009; Lang *et al.*, 2006), small molecule screen (Tanaka *et al.*, 2005), subcellular localization (Huang and Murphy, 2004; Lin *et al.*, 2007), etc. These technologies make it possible to visualize, trace and quantify cellular morphological changes in high resolution and play an increasingly crucial role to the understanding of biological processes.

Fig. 1. Representative microscopic fluorescent images of single cells with different levels of fragmentation of mitochondria. Intact mitochondria forms filamentous networks (top row), completely fragmented mitochondria have round shape (bottom row) and partially fragmented mitochondria (middle row).

Since the cost of acquiring a large number of microscopic cell images is decreasing, the analysis tools must be powerful. Consider a study aiming at correlating drug influences to morphological changes of mitochondria. This topic has become increasingly important mainly because of its relation to apoptosis and aging process. Figure 1 shows some representative cell images with different levels of fragmentation of mitochondria. As we can see, intact or completely fragmented mitochondria are relatively easy to recognize, but it is difficult to quantify partial fragmentation. In fact, in addition to fragmentation, mitochondria may become tubular, shortened, elongated or swollen. All these morphological changes are important biological indicators (Brooks *et al.*, 2007; Lee *et al.*, 2004; Tamai *et al.*, 2008). However, off-the-shelf analysis software only distinguishes morphological patterns at cellular level. No partial change or mixture of patterns at subcellular level can be readily detected. Quantification requires human inspection, which is infeasible for high-throughput screening (Taguchi *et al.*, 2007; Zhou and Wong, 2006).

Our goal is to derive an objective and reliable quantification tool to correlate drug influences and morphological differences at cellular or subcellular levels. We assume that cell images are characterized

---

*To whom correspondence should be addressed.
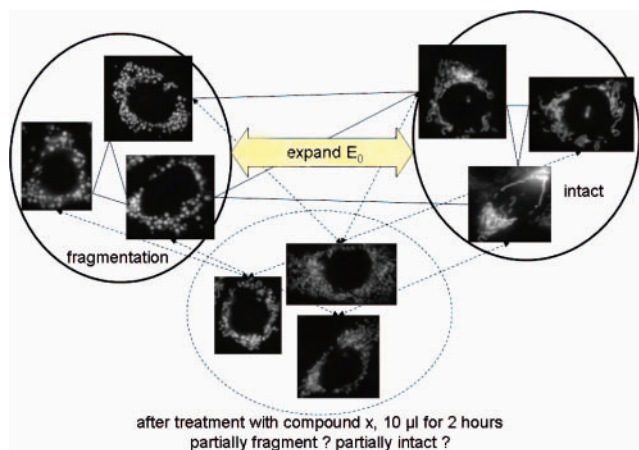†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First authors.

Fig. 2. Given two sets of cell images with extremely different morphology patterns as training examples; our method transforms the feature space so that regularized graph transition energy between the two sets is minimized. Then we can quantify the morphological difference of cells by computing their graph transition energy with the training examples in the transformed feature space.

by a feature set extracted from the images and can be considered as a data point in the feature space. A straightforward quantification is to measure the Euclidean distance between data points or apply other sophisticated similarity metrics. However, there are key challenges that must be addressed.

- A wide variety of morphological categories must be considered.

- Influence of a drug is a random variable, the difference must be estimated between pairs of cell-image *collections*. Averaging pairwise differences of single images is not appropriate here because of the presence of noise and outliers.

- Whether a selected similarity metric can actually quantify the difference that matches the purposes of the study depends heavily on the selected features.

- *Calibration* is essential for a practical quantification method because we need to measure the confidence of the quantification.

Our solution is a graph-based approach. Previously, graph-based approaches have been shown to be effective for clustering, semi-supervised learning and image segmentation (Belkin and Niyogi, 2003b; Zhu and Ghahramani, 2002; Zhu *et al.*, 2003). In our approach, *graph transition energy* is defined to quantify the similarity between collections of images. To accommodate a wide variety and combinations of morphological differences to be quantified, we adopt a supervised paradigm where two sets of extremely different cell images are assumed to be given as training examples. For example, to quantify mitochondrial fragmentation, sets of cell images with intact and completely fragmented mitochondria are assumed to be given. Figure 2 illustrates this graph-based approach. By applying a spectral graph theoretic regularization (Chung, 1997), we developed a method to transform the feature space based on the training examples so that regularized graph energy between data points of extremely different morphology is minimized. In this way, calibration of the quantification can be achieved. Then we can quantify a new set of cell images

by computing the graph transition energy between the set and training examples in the transformed feature space. Experimental results show that our method quantifies the morphological difference more accurately and sensitively than that without transformation. Results also show that our method outperforms competing methods, including neighbourhood component analysis (NCA; Goldberger *et al.*, 2005) and the multi-variate drug profiling method by Loo *et al.* (2007). Finally, we illustrate the utility of our method with a study of Annonaceous acetogenins and their impact to mitochondrial fragmentation. Our result reveals that squamocin induces more fragmented mitochondria than muricin A.

The remainder of this article is organized as follows. Section 2 reviews related work. Section 3 presents our method. Section 4 reports experimental evaluation of our method and an application to the study of Annonaceous acetogenins. The last section concludes and discusses our future work.

2 REVIEW OF PREVIOUS WORK

We briefly review previous work on image-based approaches for drug screening, mitochondrial fragmentation, graph-based approaches and feature space transformation.

2.1 Image-based approach for drug screening

Recently, high-throughput image-based approaches have received great attention for drug screening (Carpenter, 2007; Lang *et al.*, 2006; Loo *et al.*, 2007). Among them, Loo *et al.* (2007) proposed an image-based multivariate profiling method for drug screening. In their method, support vector machines (SVMs) are applied to establish a hyperplane in the feature space between cell images representing control and images of cells under different perturbation. Then the unit normal vector of the hyperplane is used as a multivariate profile to indicate the phenotypic direction of the greatest separation between the two cellular populations. By clustering these profiles, a new compound with unknown effect can be associated with compounds with known effects to infer the effects of the new compound. Clustering these profiles may also reveal new usage of an old drug if its profile is similar to another drug with target effects. Their method is interesting in that it can consider a large number of features and is not study-dependent. However, their approach fails when the images in the feature space are not linearly separable. Also, the images may distribute in the feature space so widely that the hyperplane may have a very high sampling variance. In this article, we report an experimental comparison of our graph-based method with their approach in Section 4.2.

2.2 Mitochondrial fragmentation

Mitochondria are essential organelles in eukaryotic cells, because they play a central role in energy metabolism and the integration of diverse apoptotic stimuli. Recent investigations reveal that dynamics of mitochondria morphology is important during apoptosis as the organelle undergoes massive fragmentation. Yaffe (1999) suggested that over fragmentation of mitochondria may result in breakage of mtDNA that causes excess free radicals due to impaired functions in respiration. Other reports show that mitochondrial fragmentation is an early step of apoptosis (Desagher and Martinou, 2000; Frank *et al.*, 2001; Karbowski and Youle, 2003). However, mechanisms behind the relationship between mitochondrial fragmentation and

apoptosis is still unclear (Jeong and Seol, 2008). High-content screening can play a key role in revealing their relationship and shed new light on the discovery of potential drugs for the treatment of related diseases, such as neurodegenerative diseases and diabetes.

2.3 Graph-based approach

Graph-based approaches have been successfully applied in various machine-learning problems including classification (Belkin and Niyogi, 2003b; Goldberger *et al.*, 2005; Zhan *et al.*, 2009; Zhu *et al.*, 2003), spectral clustering (Azran and Ghahramani, 2006) and dimension reduction (Belkin and Niyogi, 2003a; Roweis and Saul, 2000; Tenenbaum *et al.*, 2000). These methods create a graph whose nodes correspond to data points, while the edge weights encode the similarity between each pair of data points. It is crucial to choose a proper distance metric to estimate the similarity as the performance of graph-based models depends considerably on the similarity measure of the graph.

Zhu *et al.* (2003) uses a Gaussian random field model to construct a weighted graph representing labeled and unlabeled data for semi-supervised learning. For dimension reduction, Isomap (Tenenbaum *et al.*, 2000) applies geodesic distance to create a graph that captures the manifold structure to recover the intrinsic dimension of the data. Though geodesic distance can serve as a useful quantification of a given type of morphological difference, the isomap method requires input data uniformly distributed in the feature space with a high density to accurately recover the manifold, a requirement that is infeasible in most of biological study because the distribution of cell images is usually very skew.

Other promising applications of the graph-based approaches in bioinformatics include immune cells detection (Chang *et al.*, 2008), cell images segmentation (Isfahani *et al.*, 2008) and protein function prediction (Borgwardt *et al.*, 2005).

2.4 Feature space transformation

Feature space transformation is widely applied to data visualization and learning algorithm enhancement. Here, we reviewed two methods, linear discriminant analysis LDA and NCA, that are closely related to our method. Fisher LDA uses the label information to find informative projection such that the separation of data of different classes can be maximized. To that end, LDA tries to maximize the *inter-class scatter matrix* and minimize the *intra-class scatter matrix* simultaneously. However, LDA suffers from a small sample size problem when dealing with high-dimensional data, the intra-class scatter matrix can be nearly singular (Chen *et al.*, 2000). In our approach, we re-weight the feature space and make no attempt to compute projections as in LDA. Therefore, the problem is not an issue here because no inverse matrix is necessary. Also, LDA is designed for classification. Quantifying difference between cell images is more of a regression problem than classification. NCA (Goldberger *et al.*, 2005) learns a linear transformation of the feature space to optimize leave-one-out (LOO) performance on training data for *k*-nearest neighbor (*k*-NN) classifiers. NCA is a non-parametric learning method and makes no assumptions about data distributions. Promising performance of NCA have been shown in many applications including face recognition (Butman and Goldberger, 2008) and hyperspectral classification (Weizman and Goldberger, 2007). Since our quantification method is based on a

distance-weighted *k*-NN, we empirically compared our method with NCA and report the results in Section 4.1.2.naveenh

3 METHODS

Our task is to re-weight the feature space to separate the data points with different labels as far apart as possible. We consider a graph-based model to solve our problem.

3.1 Adjacency graph

Let $\{(x_i,y_i)\} \in X \times Y$, $i=1,\ldots,n$ be a set of data points where $X \in \mathbb{R}^D$ is an arbitrary feature space and Y a finite set of labels. In a graph-based model, a graph $G:=\{V,E\}$ is constructed for the data points, where V corresponds to data points and $e_{ij} \in E$ connects x_i and x_j. A non-negative weight is assigned to e_{ij} to build a weight matrix $\mathbf{W}$ by a radial basis function such that:

$$w_{ij}:=\exp(-\sum_{d=1}^{D}\frac{(x_{id}-x_{jd})^2}{\sigma_d^2}), \tag{1}$$

where $\sigma_d \in \sigma$ is the length scale for the d-th coordinate in the feature space. A common practice in graph-based methods is to consider weights of *nearby* data points only. Other links will be assigned zero weights. Given the weight matrix $\mathbf{W}$, the *graph energy* (Zhu and Ghahramani, 2002) can be defined as a function to measure the stability of a set of data points:

$$\frac{1}{n}\sum_{i,j} w_{ij}I(y_i \neq y_j), \tag{2}$$

where $I(\phi)$ denotes the indicator function whose value is 1 if statement ϕ is true and 0 otherwise.

3.2 Cheeger's constant

Our goal is to transform the feature space by re-weighting σ so that the Euclidean distance in the transformed space correlates with the difference of the labels of data points. The transformation requires a set of training data points with different labels. The idea is to re-scale σ_d such that the energy is minimized. To avoid overfitting, regularization is required. Well-known regularization includes *graph Laplacian* (Belkin and Niyogi, 2003b) and *Cheeger's constant* (Chang *et al.*, 2008). It has been shown that Cheeger's constant effectively removes noise, like Laplacian, but preserves better sharp boundaries needed for classification (Chang and Moura, 2007; Chung, 1997). Therefore, we chose Cheeger regularization for our problem.

Let $\mathbf{f}:X \to Y$ be a function that labels data points in G. Cheeger regularization is defined as

$$r(\mathbf{f}):=\mathbf{f}^T\mathbf{L}\mathbf{f} - \beta\mathbf{f}^T\mathbf{D}\mathbf{1}, \tag{3}$$

where $\mathbf{D}=\mathrm{diag}(d_i)$ is a diagonal matrix with $d_i=\sum_j w_{ij}$, $\mathbf{L}=\mathbf{D}-\mathbf{W}$ is the *Laplacian* of G, $\mathbf{1}=(1,\ldots,1)^T$ and β is the non-negative weight. We note that $\mathbf{f}^T\mathbf{L}\mathbf{f}$ is the graph Laplacian regularization given in (Belkin and Niyogi, 2003b). In our problem, $\mathbf{f}$ is simply the known label assignment $\mathbf{y}$ of training data. Mapping two labels in the training data to $\{0,1\}$, we have

$$\mathbf{f}^T\mathbf{L}\mathbf{f} = \mathbf{y}^T\mathbf{L}\mathbf{y} = \frac{1}{2}(2\mathbf{y}^T\mathbf{D}\mathbf{y} - 2\mathbf{y}^T\mathbf{W}\mathbf{y})$$

$$= \frac{1}{2}\sum_{i,j} w_{ij}(y_i-y_j)^2 = \frac{1}{2}\sum_{i,j} w_{ij}I(y_i \neq y_j).$$

To extend its applicability to multiple label problems, we modified the second term $\mathbf{f}^T\mathbf{D}\mathbf{1}$ in Equation (3) so that we only count if the labels are the same or not. That is, $\frac{1}{2}\sum_{ij} w_{ij}I(y_i \neq y_j) + \sum_{ij} w_{ij}I(y_i = y_j)$. Therefore, the Cheeger regularization is modified as

$$\frac{1}{2}\sum_{ij} w_{ij}I(y_i \neq y_j) - \lambda \cdot \frac{1}{2}\sum_{ij} w_{ij}I(y_i = y_j). \tag{4}$$

3.3 Graph transition energy

We embed the constraint that scaling factors sum to one for each i into Equation (4) by replacing the weight matrix $\mathbf{W}$ with a transition matrix $\mathbf{T}$ where $T_{ij} := \frac{w_{ij}}{\sum_k w_{ik}}$. With the normalization term, $\mathbf{T}$ is usually asymmetric. By doing so, we also need to modify the definition of graph energy function given in Equation (2) by replacing $\mathbf{W}$ with $\mathbf{T}$, which leads to the *Graph Transition Energy*, defined as a function of σ:

$$E(\sigma) := \frac{1}{n} \sum_{i,j} T_{ij} I(y_i \neq y_j). \quad (5)$$

A low E indicates that G is stable because data points with different labels are far apart, and vice versa. Therefore, given two sets of cell images as feature vectors, if their graph transition energy is low than they are quite different, and vice versa.

3.4 Feature space transformation

We continue our derivation of an objective function from Cheeger's constant. In addition to embedding the constraint to the objective function, we also re-parameterize $\frac{1}{\sigma_d^2} = e^{\kappa_d}$ in Equation (1) to enforce another constraint so that $\sigma_d \geq 0, d = 1, \ldots, D$. Then

$$w_{ij} = \exp\{-\sum_{d=1}^{D} e^{\kappa_d}(x_{id} - x_{jd})^2\}.$$

Let κ be a vector with κ_d's as its components. We complete the derivation of our objective function as a function of κ:

$$J(\kappa) = \sum_{ij} T_{ij} I(y_i \neq y_j) - \lambda \cdot \sum_{ij} T_{ij} I(y_i = y_j). \quad (6)$$

Minimizing J creates a transformed feature space with wide label separation and therefore calibrates the quantification with the graph transition energy. Any minimization method can be applied to minimize J. We simply used a steepest gradient descent method in our implementation.

The gradient for the objective function is

$$\frac{\partial J}{\partial \kappa_d} = \sum_{ij} I(y_i \neq y_j) \frac{\partial T_{ij}}{\partial \kappa_d} - \lambda \cdot \sum_{ij} I(y_i = y_j) \frac{\partial T_{ij}}{\partial \kappa_d}. \quad (7)$$

In Equation (7), the computation of $\frac{\partial T_{ij}}{\partial \kappa_d}$ by using chain rule is

$$\frac{\partial T_{ij}}{\partial \kappa_d} = \sum_{s,t} (\frac{\partial T_{ij}}{\partial w_{st}} \cdot \frac{\partial w_{st}}{\partial \kappa_d}), \quad (8)$$

where $\frac{\partial T_{ij}}{\partial w_{st}}$ is

$$\frac{\partial T_{ij}}{\partial w_{st}} = \frac{I(s=i)}{\sum_k w_{ik}}[I(t=j) - T_{ij}] \quad (9)$$

and $\frac{\partial w_{st}}{\partial \kappa_d}$ is

$$\frac{\partial w_{st}}{\partial \kappa_d} = -e^{\kappa_d} w_{ij}(x_{id} - x_{jd})^2. \quad (10)$$

Substituting Equation (9) and (10) into (8), we obtain

$$\frac{\partial T_{ij}}{\partial \kappa_d} = \frac{e^{\kappa_d}}{\sum_k w_{ik}}[T_{ij}(\sum_t w_{it}(x_{id} - x_{td})^2) - w_{ij}(x_{id} - x_{jd})^2]. \quad (11)$$

Recall that each data point in the graph only connects to its nearest neighbors. This approximation is sufficient because in a Gaussian energy model, links of points too far apart are nearly zero. But we cannot rule out the possibility that a pair of points becomes closer with a new κ after optimization. Therefore, we design an iteration method where each iteration applies a gradient descent update to obtain a new κ. Then the graph is rewired by connecting each point with its nearest neighbors after each iteration. The iteration repeats until convergence.

We can derive a lower bound of the objective function as follows.

PROPOSITION 3.1. *If there exists a κ such that $J(\kappa) = -\lambda \cdot n$, where n is the data size, then the data in the transformed feature space can be separated completely. That is, the transition probability $\Pr(x_i \to x_j) \geq 0$ if $y_i = y_j$, otherwise $\Pr(x_i \to x_j) = 0$*

PROOF. Because under perfect label separation, we have $T_{ij} \geq 0$ if $y_i = y_j$ and 0 otherwise. In that case $J(\kappa) \geq -\lambda \cdot n$. ∎

According to Proposition 3.1, the value of $J(\kappa)$ reveals useful information to determine if the data with different labels or conditions can be well-separated or not given the set of features. If $J(\kappa)$ is still far away from its lower bound after the re-weighting process, the feature set may not be sufficient or appropriate for representing the data points. The re-weighted κ also provides useful information to rank important of the features.

3.5 Quantifying differences

We summarize our method as follows:

(1) Define the label space Y. For example, we may define Y as different levels of mitochondrial fragmentation, days after drug applications or drug dosages.

(2) Acquire data points (i.e. cell images) with labels of the two extremely different cases (e.g. mitochondrial intact and complete fragmentation).

(3) Extract features for X and minimize Equation (6) to obtain a transformed feature space.

(4) Then we can acquire more data points and apply Equation (5) to quantify the difference.

3.6 Feature extraction

We extracted two types of features from each single cell image to serve as the input in our experiments. Strong detectors are knowledge-driven features that are supposed to provide strong hints, while weak detectors are randomly extracted patterns to allow the learning algorithm to consider subtle characteristics of a class. This set of features has been used in (Lin *et al.*, 2007) on recognizing fluorescent protein-tagged subcellular organelles in cell images, including mitochondria.

3.6.1 Strong detectors Our strong detectors consist of both geometric and texture features. Table S1 in the supplementary data gives the list of geometric features. The texture features are extracted based on the gray-level co-occurrence matrix (GLCM) method proposed by Haralick (1979). Twelve GLCM with distances 1, 2 and 10 and angles $0°, 45°, 90°$ and $135°$ are applied to the bi-leveled images. Then, various co-occurrence quantities including entropy, energy, contrast, homogeneity and correlation can be evaluated from the co-occurrence matrix to produce our texture features. The definition of these quantities are given in Supplementary Table S2.

The above method extracts a total of 155 geometric features and 500 texture features from a cell image. We conducted a backward stepwise discriminant analysis to select 132 most discriminant features as our final set of strong detectors.

3.6.2 Weak detectors The weak detectors for each image are extracted in four steps as follows.

(1) Randomly pick five images of each class as templates.

(2) Randomly extract a set of eight fragments from each template. The fragments vary in size from 9×9 to 25×25.

(3) Convolve each fragment i with a set of four filters. This set includes the original image, x derivative, y derivative and a Laplacian filter.

(4) Then for a given image either for training or testing, apply normalized cross-correlation between the given image and the fragment i to find where the fragment i occurs and then record the maximum correlation
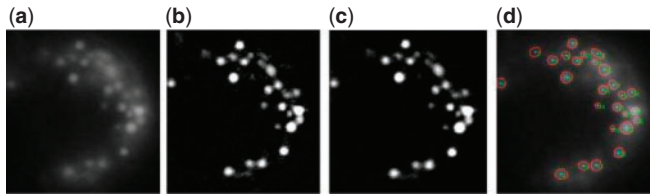
Fig. 3. The results of Mito-Q image processing. (**a**) The original image; (**b**) the image processed by the morphological filter; (**c**) the image after power-law transformation; (**d**) the locations of mitochondrial identified by Mito-Q.

as the i-th component in the vector of weak detectors for the given image.

In this study, we have two extreme classes of mitochondrial images: intact and fragmented completely. We can generate 5(template) $\times$ 8(*fragment*) $\times$ 4(filter) $\times$ 2(class)$=$ 320 weak detectors for each image. If we want to generate more features, we just increase the number of templates, fragments or filters.

3.7 Mito-Q

Because we applied our method to detect the mitochondrial fragmentation in microscopy cell images, we have developed an auxiliary tool called *Mito-Q* to assist human inspectors to visually quantify the percentage of fragmentation of mitochondria for each cell image as our golden standard. Mito-Q helps human inspectors to filter out noise and poorly focused light spots in an input image and identify which patterns correspond to mitochondria. A morphological filter (Serra, 1998), with disk structuring element (of size 2), and power-law transformation with the basic form $s=cr^{\gamma}$ (Gonzalez and Woods, 2002) are used to enhance the mitochondria objects. A bi-level operator with an optimal threshold T_H, selected by the human inspector, is applied to generate the template of mitochondrial, as shown in Figure 3. At last, Mito-Q will generate mitochondrial features for evaluating the percentage of fragmentation. Though Mito-Q is helpful, it takes a skillful human inspector to generate desirable output scores and is not designed to replace automated quantification methods.

4 RESULTS AND DISCUSSION

This section reports the empirical evaluation of our feature space transformation method.

4.1 Effectiveness of feature space transformation

We first evaluate whether our graph-based feature space transformation method actually results in a new space that characterizes the difference between data points better. To this end, we evaluate whether our method improves a semi-supervised learning task and a k-NN regression task, both for cell-image analysis.

4.1.1 Improving semi-supervised learning We applied our feature space transformation method to improve the classification performance of Zhu *et al.*'s (2003) semi-supervised learning method for the task of recognizing subcellular organelle patterns in the HeLa cell image data set. We used the 2D HeLa image data set available from the Murphy lab (See http://murphylab .web.cmu.edu/data/2Dhela_images.html). This data set contains 862 images. Each image contains a single cell with exactly one of the 10 distinct subcellular organelles tagged by a fluorescent
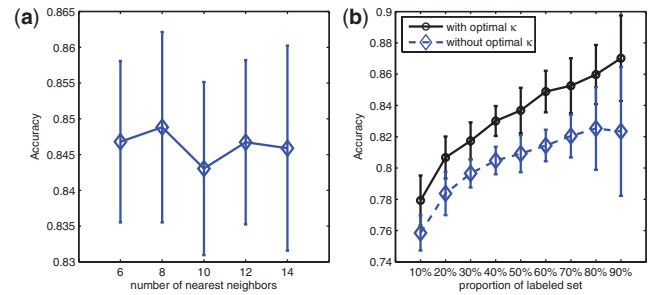
Fig. 4. (**a**) Influence of the k nearest neighbors. (**b**) The performance of semi-supervised learning before and after learning κ in the HeLa cell image data set.

protein. The feature set that we used was the superset of SLF16, which consists of 47 features selected by stepwise discriminant analysis (SDA) from a set of 180 features (See http://murphylab.web.cmu.edu/data/2Dhela_images_download.html). Among them, six are DNA-channel features. We used this 180-feature set of the HeLa images in our first experiment.

We split the test images into two subsets. One of the subsets was used as the labeled data and the other as unlabeled data. When our method was applied, the labeled data were also used to transform the feature space. We must first determine k, the number of the nearest neighbors to construct a graph. We also wanted to see if our method is sensitive to k. We selected the k-value from $\{6,8,10,12,14\}$ according to their classification performance. We repeated the experiments on each data set for 10 runs with random partitions of 60% training/40% test instances. It can be observed from Figure 4a that there is no significant difference for different k. Therefore, we arbitrarily assigned $k=8$, the best performing setting, to construct the graph in our experiments. We also found that our method is insensitive to choices of λ in Equation (6) over a wide range. In all of our experiments, we simply set $\lambda=0.1$.

The result of semi-supervised learning is given in Figure 4b, which shows that though both semi-supervised methods improve their performance as the number of labeled training examples increases, applying our transformation clearly improves the performance of the plain semi-supervised learning method. The best accuracy is about 87%, about as good as a supervised neural net reported in (Huang and Murphy, 2004) for the same set of images. We expect that with more unlabeled images, semi-supervised learning with transformed feature space has the potential to achieve an accuracy as good as the best results by supervised learning.

Moreover, we can compute graph transition energy between images of different subcellular organelles. The resulting energy table can serve as a measure of pairwise similarity. Then we can applied multi-dimensional scaling (MDS) to help us visualize the morphological similarity of the shape of 10 distinct subcellular organelles as shown in Figure 5, where we can see that two Golgi types (Gpp130 and Giantin) are particularly challenging to distinguish, so are mitochondria and tubulin.

4.1.2 Estimating partial fragmentation of mitochondria Now we show that our approach can also be applied to improve the estimation in the fragmentation level of mitochondria in a single cell without actually identifying each mitochondrion object in a cell image. Since estimating fragmentation of a mitochondrion is
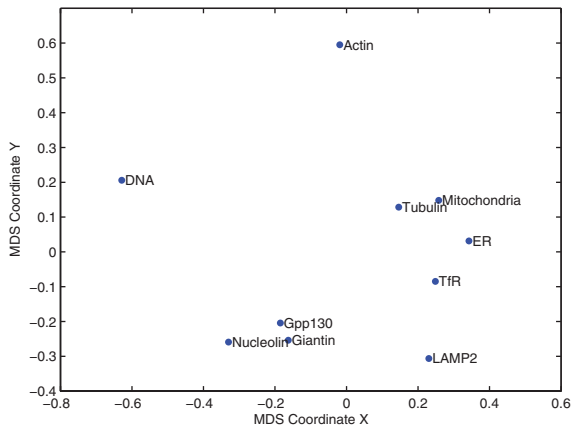
Fig. 5. Visualizing shape similarity of 10 distinct subcellular organelles with a MDS plot derived from pair-wise graph transition energy.

Table 1. Comparing the performance of estimating partial fragmentation of mitochondria in individual cell images

	k-NN (before learning κ)	NCA	k-NN (after learning κ)
Correlation	0.696	0.741	0.814
MSE	1.610	1.409	1.006

a regression problem, we applied distance weighted k-NN for this task. We compared our feature space transformation method with NCA (Goldberger *et al.*, 2005), which is designed to transform the feature space to minimize the error of k-NN.

To apply our method, in addition to training data for obtaining an optimal κ, we need other labeled data in the feature space to approximate the manifold of the data. With the labeled data, we can quantify each test image by the weighted average of its nearest neighbors:

$$\hat{f}_i = \frac{\sum_{k=1}^{K}(w_{ik} * f_k)}{\sum_{k=1}^{K}(w_{ik})}, \tag{12}$$

where f_k is the label of the k-th labeled data and K is a pre-defined constant for the number of the nearest neighbors. In our test case, f_k is simply the score by human inspectors with aid of Mito-Q and $K = 6$. We used the same κ obtained from mostly impact (MI) and mostly fragmented completely (MC) and all 392 images as the labeled data. We then acquired an additional set of 43 new images to evaluate the above method.

Table 1 shows the performance comparison in terms of correlation coefficient with the human inspectors' score and mean-square-error (MSE). The result shows that our method outperforms NCA by achieving the highest correlation coefficient and the lowest MSE. The result also shows that the transformation by learning κ effectively improves the quality of the estimation.

4.2 Quantifying collective morphological difference

We now consider the main task that our approach is designed for. That is to quantify morphological difference between collections of cell images. This task is particularly useful for drug screening, as
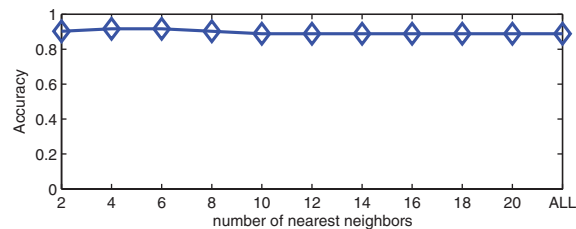

Fig. 6. Influence of the k-NN.

different image sets represent the results of the treatment of different drugs. The same approach can be applied to study effects of the same drug under different dosages and time courses.

To test our method for this task, we acquired a collection of 392 single cell images with different percentage of mitochondria fragmentation. We sorted the cell images according to the output scores of Mito-Q and labeled the top 36 images as MI and bottom 36 as MC to construct two extreme cases of mitochondria. The rest of the images (320) are labeled as mitochondrial fragmented partially (MP) in three equal sized subsets (MP_1, MP_2 and MP_3), from the least fragmented to the most. For each cell image, we extracted 452 features that include 132 strong (e.g. morphology, texture, moment and intensity features) and 320 weak detectors, as described in Section 3.6.

We sampled with replacement from MI and MC to create 11 treatment sets with different mixture proportion. The k-th set was a mixture of $(11-k) \times 10\%$ MI images with $(k-1) \times 10\%$ MC. Then we regarded raw MI as the control set to compute the difference between each mixture set and control set. Note that the percentage here refers to the proportion of images sampled from MI or MC, not the degree of mitochondrial fragmentation. We compared our approach with the approach proposed by Loo *et al.* (2007). Recall that in their approach, SVM is applied to establish a hyperplane in the feature space to distinguish images of cells under different perturbation. Then the unit normal vector of the hyperplane is used as a profile and cosine similarity between profiles is used to quantify the difference. To duplicate their method, we obtain a profile between control and each treatment set. In our graph-based method, we use graph transition energy defined in Equation (5) to quantify the difference in the transformed feature space by minimizing Equation (6) with MI and MC as the training examples. Because of a small set of training data, we determine the number of the nearest neighbors from $\{2, 4, 6, \ldots, 20, \infty\}$ via leave-one-out validation based on the classification performance of Zhu *et al.*'s (2003) semi-supervised learning method. It is observed from Figure 6 that the accuracy rate is robust to the number of the nearest neighbors. Therefore, we just use six nearest neighbors for this task.

We repeated the above trial 1000 times to approximate the distribution of the similarity of profiles and the distribution of the energy values to determine which method has higher power of discrimination. In our experiment, we also compared the energy distribution obtained in the feature space without transformation to demonstrate the effectiveness of the re-weighting process. Figure 7 plots the distribution of the resulting 1000 quantification for the 1st (pure MI, 0% MC), 6th (mixture of MI and MC with equal proportions, 50% MC) and 11th (pure MC, 100% MC) treatment sets by different approaches. The result shows that the approach proposed by Loo *et al.* (2007) yields a large overlapping area among
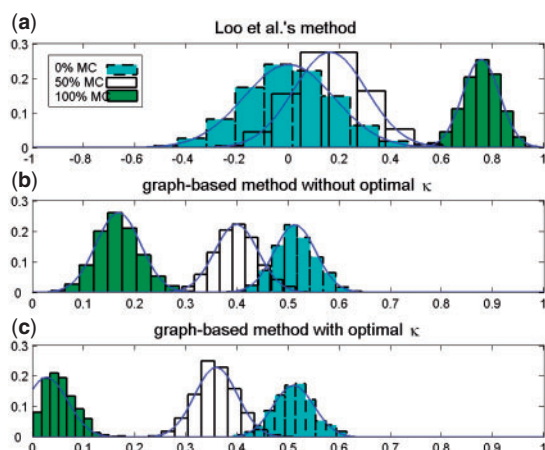
Fig. 7. Distributions of the similarity of profiles as proposed by Loo *et al.* (2007) (**a**) and distributions of energy estimated by the graph-based method with (**c**) and without (**b**) feature space transformation for the 1st (pure MI, 0% MC), 6th (mixture of MI and MC with equal proportions, 50% MC) and 11th (pure MC, 100% MC) treatment sets.
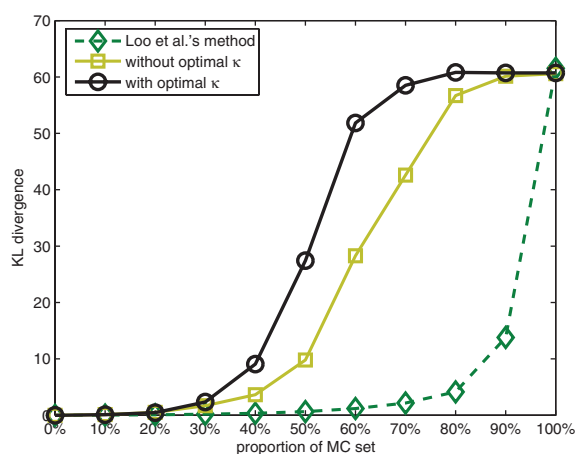
Fig. 8. KL divergence of the distributions of the quantification of the difference between control (pure MI) and the treatment sets with increasing proportion of MC.

these distributions, implying that it is quite likely that their approach will fail to quantify that a sub-sample of MI is more similar to raw MI (control) than a mixture set of MI and MC. In contrast, our method yields a small overlapping area and will quantify the difference in a correct order. Though without transformation, the overlapping area is still small, but the mean between pure MI and pure MC is closer than that with transformation. Next, for all treatment sets, we estimate the KL divergence to measure the difference of the distributions of the quantification. For an ideal method, the KL divergence should increase gradually and then surge immediately after the proportion of MC is increased to more than 50%. Figure 8 plots the result, which reveals that our approach has the highest discrimination power and that the transformation of the feature space actually improves the quality of quantification substantially.

Next, we evaluated whether the quantification by our approach actually reflects the morphological difference. We computed the

Table 2. Graph transition energy between the treatment sets with different levels of mitochondrial fragmentation

	MI	MP$_1$	MP$_2$	MP$_3$	MC
MI		0.321	0.269	0.164	0.024
MC	0.024	0.092	0.145	0.261	
Mito-Q range	0.00–0.05	0.05–0.15	0.15–0.29	0.29–0.79	0.82–1

energy between each partially fragmented subset and MI/MC in the transformed feature space. The result is shown in Table 2. As expected, the energy values are in the same order as their score of fragmentation measured by human inspectors with aid of Mito-Q. In our approach, the value of the objective function $J(\kappa)$ indicates how well the data are separated in the transformed feature space. If $J(\kappa)$ approaches the lower bound of the objective function given in Proposition 3.1, then the data of extremely different classes can be separated well, which implies that the feature set characterizes the difference well, and that the confidence of our quantification is high. In this case, the lower bound of J is equal to -7.2 ($\lambda = 0.1$ and $n = 72$). With the optimal κ learned from MI and MC sets, $J(\kappa)$ can reach -6.040 in the transformed space, suggesting a high confidence for the quantification.

4.3 Analysis of *Annonaceous acetogenins*

We applied our method to the study of Annonaceous acetogenins. More than 250 compounds in the Annonaceous acetogenin family have been isolated from *Annonaceae* plants. In this article, we report our results with muricin A and squamocin, two members in the family with the potential of triggering mitochondria fragmentation and apoptosis in tumor cells. We measured the effects of the compounds at different time points by quantifying mitochondrial morphology changes on Chinese hamster ovary (CHO) cells for 3 days. Besides, we consider the effect of DMSO as the control. For each group (a compound–day combination), we acquired images, segmented them into 20–70 single cells, and then applied our method to analyze the results. We found that the values of graph transition energy measured against DMSO for both muricin A and squamocin decrease over time as shown in Figure 9. The lower the energy the higher the morphological difference compared with the DMSO groups. That is, from Figure 9, we can conclude that squamocin induces more fragmented mitochondria than muricin A in CHO cells. This conclusion can be verified by human inspection. Figure 10 shows the representative images in each group. We can clearly see that on day 2, mitochondria of squamocin-treated cells have large round shape, indicating a high degree of fragmentation, while on the same day, mitochondria of muricin A-treated cells are much smaller. Our approach allows us to automatically quantify to what extent the difference is. This is crucial for large scale studies where human inspection becomes increasingly time-consuming and error-prone.

Again, we can compute graph transition energy between each pair of image groups of muricin A and squamocin and apply MDS to plot all image groups in a 2D space as shown in Figure 11. Arrows in the plot indicate the time course of the morphological change by these compounds. The plot clearly shows a drastic change of mitochondrial morphology from day 1 to day 2 in squamocin-treated cells.
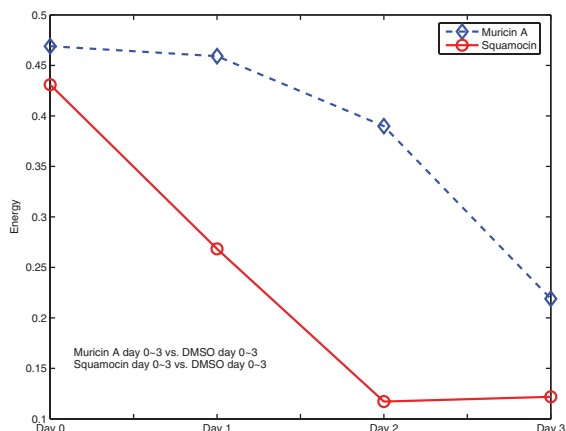
Fig. 9. Comparing the effect of mitochondrial fragmentation by squamocin and muricin A in CHO cells.


Day 0   Day 1   Day 2   Day 3
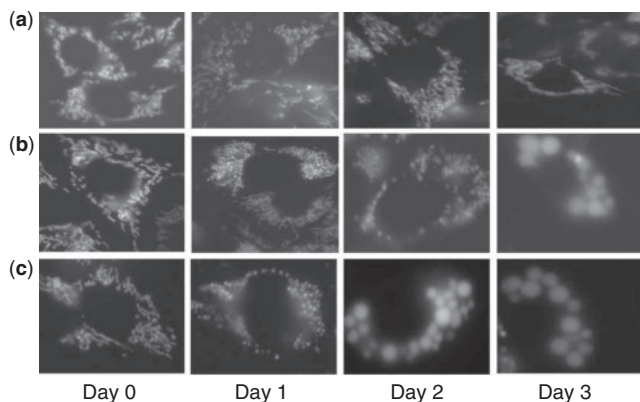
Fig. 10. Representative images of the cells after treated by DMSO (**a**), muricin A (**b**) and squamocin (**c**) for a certain number of days.

5 CONCLUSION

In the end, we summarize our contributions as follows:

- We defined *graph transition energy* to quantify morphological difference of two cell-image collections under different perturbations.

- We applied a spectral graph theoretic regularization to re-weight the feature space according to training examples of extremely different cases. Experimental results show that our feature space transformation method actually improves the quality of classification and quantification. The method also allows for calibration by providing a confidence score of the quantification.

- We demonstrated how to quantify morphological difference of collections of cell images in our approach and showed that our approach has a higher discrimination power than a competing approach by Loo *et al.* (2007).

- We showed that after applying our feature space transformation method, the performance of a semi-supervised learning method by Zhu *et al.* (2003) for subcellular localization can be improved significantly.
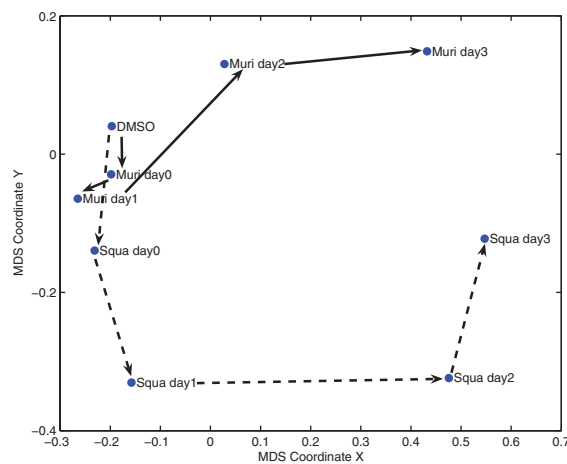
Fig. 11. Morphological similarity of mitochondria in different image groups by MDS according to the pairwise graph transition energy values.

- Coupling with a *k*-NN regression method, we demonstrated that our transformation method improves the accuracy of the estimation of partial fragmentation of mitochondria in cell images, and that our method is more effective than NCA.

- In Section 4.3, we demonstrated a case study of two Annonaceous acetogenin compounds and their effects to mitochondrial fragmentation.

Our approach is not specific to any particular morphological differences and can potentially be applied to quantify any difference with an appropriate set of features. Therefore, it has the potential to rapidly screen many drug candidates and understand their effects. In our future work, we will apply this approach to continue our study of Annonaceous acetogenins and their various effects to mitochondria.

REFERENCES

Azran,A. and Ghahramani,Z. (2006) A new approach to data driven clustering. In *ICML '06: Proceedings of the 23rd international conference on Machine learning*. ACM, New York, NY, USA, pp. 57–64.

Belkin,M. and Niyogi,P. (2003a) Laplacian eigenmaps for dimensionality reduction and data representation. *Neur. Comput.*, **15**, 1373–1396.

Belkin,M. and Niyogi,P. (2003b) Using manifold stucture for partially labeled classification. In *Advances in Neural Information Processing Systems 15 (NIPS*2002)*. MIT Press, Cambridge, MA, pp. 929–936.

Borgwardt,K.M. *et al*. (2005) Protein function prediction via graph kernels. *Bioinformatics*, **21**(suppl. 1), 47–56.

Brooks,C. *et al*. (2007) Bak regulates mitochondrial morphology and pathology during apoptosis by interacting with mitofusins. *Proc. Natl Acad. Sci. USA*, **104**, 11649–11654.

Butman, M. and Goldberger, J. (2008) Face recognition using classification based linear projections. *EURASIP J. Adv. Signal Proc.*, **8**, 1–7.

Carpenter,A.E. (2007) Image-based chemical screening. *Nat. Chem. Biol.*, **3**, 461–465.

Chang,H.-H. and Moura,J.M.F. (2007) Classification by Cheeger constant regularization. In *Proceedings of IEEE International Conference on Image Processing (ICIP-2007)*. Volume II, IEEE Publications Office, Los Alamitos, CA. pp. II209– II212.

Chang,H.-H. *et al*. (2008) Automatic detection of regional heart rejection in USPIO-enhanced mri. *IEEE Trans. Med. Imag.*, **27**, 1095–1106.

Chen,L.-F. *et al*. (2000) A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recogn.*, **33**, 1713–1726.

Chung,F.R.K. (1997) *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics. American Mathematical Society, Providence, RI.

Desagher,S. and Martinou,J.-C. (2000) Mitochondria as the central control point of apoptosis. *Trends Cell Biol.*, **10**, 369–377.

Frank,S. *et al*. (2001) The role of dynamin-related protein 1, a mediator of mitochondrial fission, in apoptosis. *Dev. Cell*, **1**, 515–525.

Goldberger,J. *et al*. (2005) Neighbourhood components analysis. In *Advances in Neural Information Processing Systems 17 (NIPS*2004)*. MIT Press, Cambridge, MA, pp. 513–520.

Gonzalez,R.C. and Woods,R.E. (2002) *Digital Image Processing,* 2nd edn. Prentice Hall, Upper Saddle River, NJ.

Haralick,R. M. (1979) Statistical and structural approaches to texture. *Proceedings of IEEE*, **67**, 786–804.

Huang,K. and Murphy,R.F. (2004) Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. *BMC Bioinformatics*, **5**, 78.

Isfahani,S.N. *et al*. (2008) A new approach for touching cells segmentation. In *International Conference on BioMedical Engineering and Informatics*, IEEE Computer Society, Washington, DC. pp. 816–820.

Jeong,S.-Y. and Seol,D.-W. (2008) The role of mitochondria in apoptosis. *BMB Rep.*, **41**, 11–22.

Jones,T. R. *et al*. (2009) Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proc. Natl Acad. Sci. USA*, **106**, 1826–1831.

Karbowski,M. and Youle,R.J. (2003) Dynamics of mitochondrial morphology in healthy cells and during apoptosis. *Cell Death Differ.*, **10**, 870–880.

Lang,P. *et al*. (2006) Cellular imaging in drug discovery. *Nat. Rev. Drug Discove.*, **5**, 343–356.

Lee,Y.-J. *et al*. (2004) Roles of the mammalian mitochondrial fission and fusion mediators Fis1, Drp1, and Opa1 in apoptosis. *Mol. Biol. Cell*, **15**, 5001–5011.

Lin,C.-C. *et al*. (2007) Boosting multiclass learning with repeating codes and weak detectors for protein subcellular localization. *Bioinformatics*, **23**, 3374–3381.

Loo,L.-H. *et al*. (2007) Image-based multivariate profiling of drug responses from single cells. *Nat. Methods*, **4**, 445–453.

Roweis,S.T. and Saul,L.K. (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**, 2323–2326.

Serra,J. (1998) *Image Analysis and Mathematical Morphology*. Academic Press, Orlando, FL.

Taguchi,N. *et al*. (2007) Mitotic phosphorylation of dynamin-related gtpase drp1 participates in mitochondrial fission. *J. Biol. Chem.*, **282**, 11521–11529.

Tamai,S. *et al*. (2008) Characterization of the mitochondrial protein letm1, which maintains the mitochondrial tubular shapes and interacts with the aaa-atpase bcs1l. *J. Cell Sci.*, **121**, 2588–2600.

Tanaka,M. *et al*. (2005) An unbiased cell morphology-based screen for new, biologically active small molecules. *PLoS biology*, **3**.

Tenenbaum,J.B. *et al*. (2000) A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319–2323.

Weizman,L. and Goldberger,J. (2007) A classification-based linear projection of labeled hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, IEEE Publications Office, Los Alamitos, CA. pp. 3202–3205.

Yaffe,M.P. (1999) Dynamic mitochondria. *Nat. Cell Biol.*, **1**, 149–150.

Zhan,D.-C. *et al*. (2009) Learning instance specific distances using metric propagation. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, New York, NY, USA, pp. 1225–1232.

Zhou,X. and Wong,S.T. (2006) Informatics challenges of high-throughput microscopy. *IEEE Signal Proc. Mag.*, **23**, 63–72.

Zhu,X. and Ghahramani,Z. (2002) Learning from labeled and unlabeled data with label propagation. *Technical Report CMU-CALD-02-107*, Carnegie Mellon University, Pittsburg, PA.

Zhu,X. *et al*. (2003) Semi-supervised learning using gaussian fields and harmonic functions. In *ICML '03: Proceeding of the 12th International Conference on Machine Learning*. AAAI Press, Menlo Park, CA, pp. 912–919.