

RESEARCH ARTICLE

Open Access

Similarity maps and hierarchical clustering for annotating FT-IR spectral images

Qiaoyong Zhong¹, Chen Yang¹, Frederik Großerüschkamp², Angela Kallenbach-Thieltges², Peter Serocka^{1,2}, Klaus Gerwert^{1,2} and Axel Mosig^{2*}

Abstract

Background: Unsupervised segmentation of multi-spectral images plays an important role in annotating infrared microscopic images and is an essential step in label-free spectral histopathology. In this context, diverse clustering approaches have been utilized and evaluated in order to achieve segmentations of Fourier Transform Infrared (FT-IR) microscopic images that agree with histopathological characterization.

Results: We introduce so-called *interactive similarity maps* as an alternative annotation strategy for annotating infrared microscopic images. We demonstrate that segmentations obtained from interactive similarity maps lead to similarly accurate segmentations as segmentations obtained from conventionally used hierarchical clustering approaches. In order to perform this comparison on quantitative grounds, we provide a scheme that allows to identify non-horizontal cuts in dendrograms. This yields a validation scheme for hierarchical clustering approaches commonly used in infrared microscopy.

Conclusions: We demonstrate that interactive similarity maps may identify more accurate segmentations than hierarchical clustering based approaches, and thus are a viable and due to their interactive nature attractive alternative to hierarchical clustering. Our validation scheme furthermore shows that performance of hierarchical two-means is comparable to the traditionally used Ward's clustering. As the former is much more efficient in time and memory, our results suggest another less resource demanding alternative for annotating large spectral images.

Keywords: Hierarchical clustering, Cluster validation, FT-IR microscopy, Raman microscopy, Image analysis, Similarity maps

Background

In recent years, it has been well-established that label-free Fourier transform infrared (FT-IR) microscopy can resolve pathologically relevant information from histological tissue samples [1-3], as surveyed in [4]. Unveiling histopathologically relevant structures from localized absorbance spectra yielded by an FT-IR microscope, schematically illustrated in Figure 1, is typically achieved through a combination of unsupervised and supervised learning approaches [5,6]. First, certain number of spectrally measured tissue sections are being *annotated* based on pre-segmented spectral images, typically based on unsupervised clustering [6-8]. These annotations are then

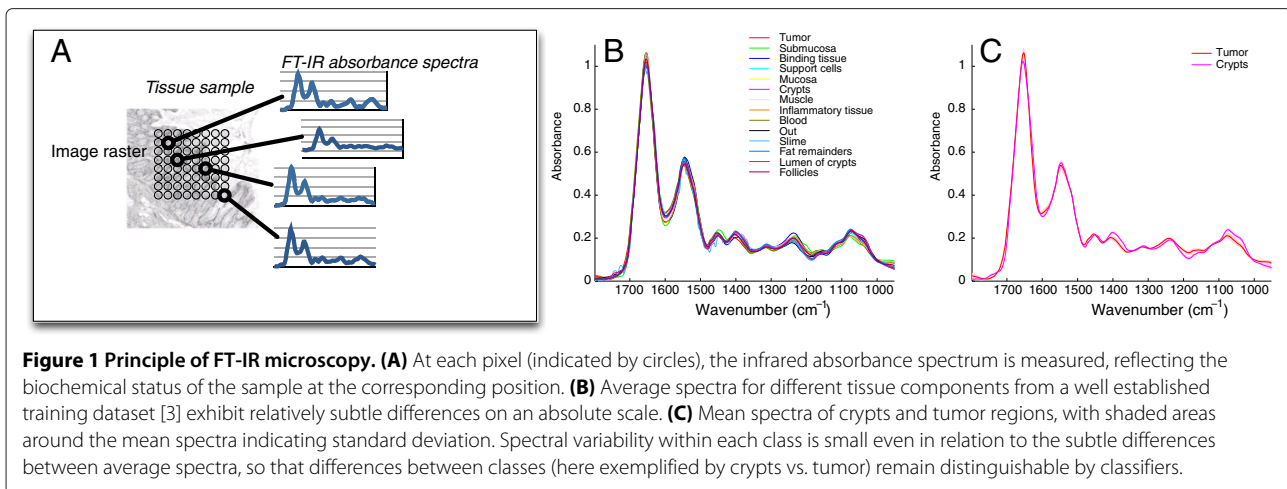
used to extract spectra as training data for supervised classifiers. Obviously, the quality of the annotation determines what tissue components can be resolved and how reliably they can be recognized by spectral classifiers. In this context, we introduce a novel interactive approach to annotation and quantitatively validate this approach in comparison to established annotation schemes. In particular, we provide novel algorithms to perform such quantitative comparison. As utilizing Raman [9] or CARS [10] microscopy often underlies the same workflow of data processing as FT-IR microscopy [11], the ideas discussed here may equally apply to these other types of label-free multispectral microscopy.

For annotating spectral images, at least two strategies are commonly employed. The first straightforward approach is to cluster all image spectra into a suitable number of k clusters. Each cluster is then identified

*Correspondence: axel.mosig@bph.rub.de

²Department of Biophysics, Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, Germany

Full list of author information is available at the end of the article



with one index color, so that a pathologist may identify regions in the corresponding index color image with tissue components. As a second and typically complementary approach, the spectral image can be overlaid with a Hematoxylin and Eosin (H&E) stained image of the same tissue region. Pathologists can identify relevant tissue components in the H&E stained image, whose location patterns can be carried to the corresponding locations in the spectral image in order to extract spectra belonging to a certain tissue component. In practice, the accuracy in overlaying H&E stained images with FT-IR spectral images is limited, e.g. due to slight distortions of the tissue during the staining procedure. Also, there are limitations in identifying and marking up precise borders between certain tissue components, so that most approaches to FT-IR based spectral histopathology combine the two approaches [2,7,8]: A presegmentation of the spectral image is overlaid with the H&E stained image of the same sample, and then clusters in the spectral image are identified with tissue components based on their overlap with relevant regions in the H&E stained image, as identified by a pathologist. In general, the relation between clusters and tissue components is not one-to-one, but one tissue component may often be associated with several clusters. Thus, the number of clusters is usually chosen relatively large, so that the image is rather oversegmented. For obtaining presegmentations, hierarchical cluster analysis (HCA) [7] as well as *k*-means or fuzzy *c*-means [2,8] are common choices.

Spectral image segmentation using similarity maps

As our main contribution, we introduce a novel interactive method for annotating FT-IR spectral images. Based on so-called similarity maps [12] and utilizing the concept of certain similarity measures between high dimensional vectors, annotations result from interactively choosing reference pixel spectra for the tissue components that can

be identified in the tissue sample. By overlaying the interactive similarity maps (ISMs) with an H&E stained reference image, this allows to interactively take into account both spectral similarity and histopathological information from the stained image. This method is implemented in our so-called *Lasagne* software that has been originally proposed [13] and implemented [14] for multi-label fluorescence microscopy, while the present contribution adapts and quantitatively validates it for the use in vibrational microspectroscopy.

Clustering and its validation for spectral image segmentation

In order to convincingly establish similarity maps as a suitable tool for infrared image annotation, it is essential to compare them to the currently predominant approach using clustering-based presegmentations. Clustering methods such as Ward's method or *k*-means have been used extensively in infrared image segmentation and compared on a qualitative level for their suitability in vibrational microspectroscopy [7]. Yet, as has been noted prominently [15], "the validation of clustering is the most difficult and frustrating part of cluster analysis". In fact, different applications and different clustering algorithms — in particular hierarchical ones — require different validation approaches. One application-specific consideration when employing and validating clustering in the context of infrared image annotation is that the correspondence between tissue components is not one-to-one, but rather one-to-many. Such considerations are rarely accounted for in existing validation procedures, and may be a reason that current comparisons of clustering approaches [7] are qualitative rather than quantitative. We address the lack of quantitative evaluation through introducing a validation scheme for hierarchical clustering based on so-called tree-assignments which were introduced in [16,17] in the context of tracking cells in live cell

imaging. A very brief and preliminary validation of Ward's clustering in FT-IR image segmentation was given in [3] and is fully detailed and systematically elaborated in the present work. In particular, we evaluate the suitability of hierarchical clustering approaches that are more efficient in terms of computational resources.

To identify a suitable validation scheme for unsupervised infrared image segmentation, it is important to consider in some more detail how clustering based segmentations are commonly used in this context. While not commonly described in detail, one typically attempts to choose a number of clusters that oversegments the image in a presegmentation. Then, the task of the human annotator is to identify each tissue component with one or several of these clusters. In some approaches, clusters may be extended or divided by navigating along the hierarchy of a given dendrogram. Using a ground truth segmentation as a reference, a validation scheme thus should aim to identify the best possible segmentation obtained from this workflow from a given clustering algorithm under realistic side assumptions.

One such side assumption is that there is a limit to the number of clusters that can be merged to represent one tissue component, which essentially represents a cognitive limit of a human annotator. Obviously, starting annotation from a crudely oversegmented image in general will allow more precise annotations. However, this comes at the cost of requiring the annotator to identify many small segments that merge into one tissue component. As there is no fixed limit to the degree of oversegmentation, we propose a validation scheme that takes the degree of oversegmentation into account as a parameter. We will refer to the degree of oversegmentation utilized during annotation as the *depth* of segmentation obtained from a dendrogram, and will introduce a validation scheme that allows to control segmentation depth through a parameter.

As surveyed in [18], a large diversity of validation measures for clustering algorithms has been proposed. Our main concern in this work is to validate clusters against a ground-truth reference segmentation, which is commonly referred to as *external validation*. While for external validation, one can principally rely on measures such as accuracy known from the validation for supervised classifiers, measures such as the Rand index and the Jaccard index [19,20], both in eventually normalized forms, are well established and commonly used. Furthermore, so-called *variation of information* [21] can be considered a well established information theoretic measure. However, these measures can only be applied to fixed segmentations, but not to dendrograms obtained from hierarchical clustering, and do not account for one-to-many relations between reference classes and clusters.

In infrared image annotation and other applications it is commonplace to use the dendrogram for obtaining a

fixed partitioning into a certain number of classes. In this context, a straightforward and widely used approach horizontally cuts dendrograms into a fixed number of clusters [22]. In other words, given a dendrogram, one identifies edges e_1, \dots, e_k so that each e_i contains a point v_i that has same distance δ from the root for all i . Now, subtrees below these k edges define a partitioning into k classes. In general, however, there are numerous non-horizontal cuts supported by the same dendrogram that yield a different partitioning into the same number of clusters, which has been considered only recently in literature [23-25].

As illustrated in Figure 2, a segmentation based on a non-horizontal cut will generally reflect tissue components much better than a horizontal cut. Thus, validating different approaches to clustering in this context should take into account such non-horizontal cuts. Our contribution elaborates an approach that allows to systematically identify such non-horizontal cuts, yielding a corresponding validation scheme for hierarchical clustering. In particular, we utilize this scheme to quantitatively compare different hierarchical clustering approaches to interactive similarity maps. An important property of our validation scheme is that it can measure validity under different depths, i.e., different degrees of initial oversegmentation.

Methods

Interactive Similarity Maps (ISMs)

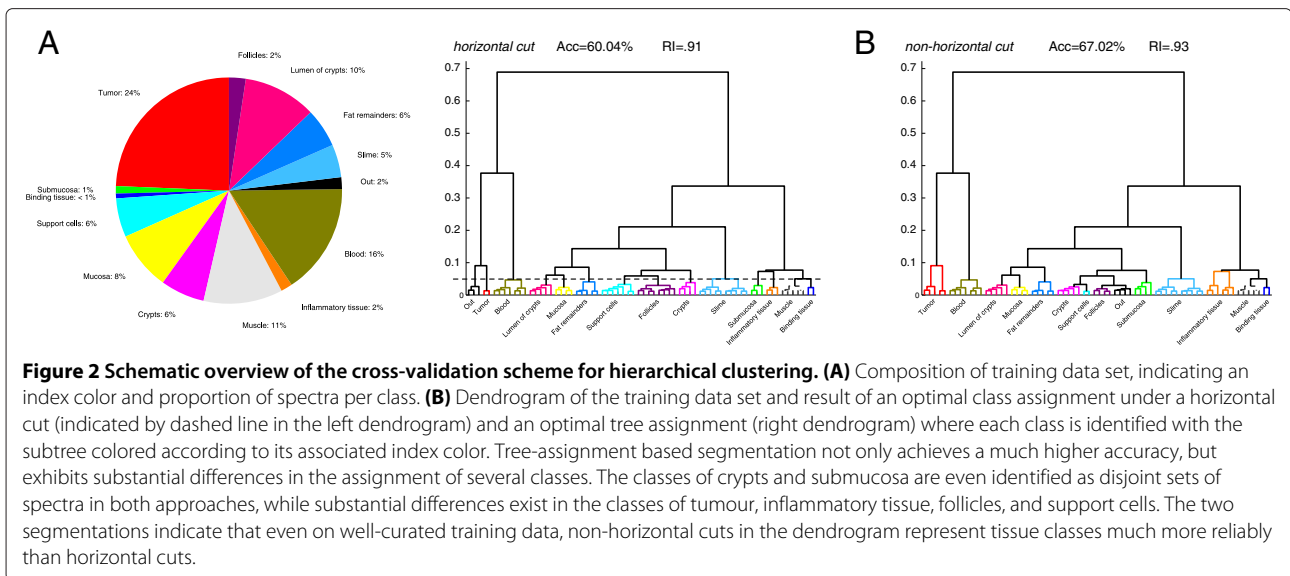
To introduce the concept of similarity maps following [12], let $F(x, y)$ denote the FT-IR absorbance spectrum at position (x, y) in the spectral image. By choosing a reference spectrum $R = F(x_R, y_R)$ at position x_R, y_R , one can measure the similarity between any position spectrum $F(x, y)$ and the reference spectrum $R = (R_1, \dots, R_n)$ using a suitable measure of spectral similarity. Now, interpreting the spectral similarity at each position as an intensity, we obtain the similarity map M as an intensity image through

$$M_R(x, y) = \sigma_R(F(x, y)),$$

where σ_R measures the spectral similarity to the reference spectrum R . We follow the suggestions from [12] in using

$$\sigma_R(S) := \prod_{i=1, \dots, n} (1 - |R_i - S_i|^\alpha) \quad (1)$$

as our similarity function between reference spectrum $R = (R_1, \dots, R_n)$ and pixel spectrum $S = (S_1, \dots, S_n)$. Here, α is a non-negative real-valued parameter to adjust the sensitivity of the similarity measure. Note that Eq. 1 only makes sense if R_i and S_i range between 0 and 1. In practice, we achieve this by rescaling a dataset, and setting the minimum absorption occurring at any wavenumber in any spectrum to 0, and correspondingly setting the maximum absorption occurring in the dataset to 1.



Interpreting $\sigma_R(S)$ as a similarity measure between vectors R and S , it has been shown that Eq. (1) satisfies metric properties, which turns out to be a metric obtained by natural and systematic scheme to induce new metrics as products of other metrics [12]. A major advantage from a practical point of view is that Eq. (1) can be implemented on graphics hardware, so that the similarity map for all spectra from one image w.r.t. a given reference spectrum R can be computed within fractions of a second. This allows an interactive exploration of a spectral image by setting and moving the coordinate for the reference spectrum R with the mouse pointer.

In general, the product in Eq. (1) may vanish towards 0 rapidly if the difference is large for only a few features. The parameter α can be used to control this effect. Large values α lessen the tendency of the product to vanish towards 0 in datasets with heterogeneous features. Choosing a small α close to 0, on the other hand, can be used to amplify the tendency for the product to vanish when working on datasets with little variability. In practice, the parameter can be adapted interactively, where common choices range between 1 and 2.

As demonstrated in Additional files 1, 2, 3, 4 and 5, an image can be annotated through similarity maps by interactively mouse-clicking one (or several) reference positions for each tissue component. This yields one reference spectrum R for each tissue component in the spectral image. Once a reference spectrum R is chosen, an intensity cutoff in the similarity map M_R can be set interactively. All positions (x, y) exceeding this threshold will be considered part of the same tissue component. If there are K different types of tissue components in the spectral image, annotation now reads as interactively identifying reference pixel spectra R_1, \dots, R_K . In practice, as discussed in

further detail in Results Section, some tissue components need to be represented by two rather than one reference spectrum. In case one position exceeds the threshold of several reference spectra, the position is assigned to the similarity map of highest intensity.

As an implementation, we utilized a version of the *Lasagne* software [13,14] adapted to the requirements of vibrational microspectroscopic data. A key feature of the *Lasagne* software is to perform computation of Eq. (1) on graphics hardware, so that the similarity maps M can be displayed in real-time. In our adapted version, the *Lasagne* software may also display overlays between the similarity maps and a reference image such as an H&E-staining image. In practice, annotation of an FT-IR microscopic dataset using the *Lasagne* software works as follows (see Additional files 1, 2, 3, 4 and 5): The annotator uploads both the FT-IR dataset and an H&E staining image of the same sample in the same coordinate system into the *Lasagne* software. Now, the spectrum at the current position of the mouse cursor in the image is interpreted as the reference spectrum to interactively display a similarity map. When moving the cursor to a suitable reference point, the similarity map may highlight a particular tissue compartment, which can be visually aligned by toggling between the similarity map and the H&E image. Once the correspondence between the similarity map and the tissue structure identifiable from the H&E image is visually well matched by varying the reference point, the annotator sets a suitable intensity threshold in the similarity map, so that the above-threshold positions can be used as training spectra for the corresponding tissue compartment. This may not only be done for one tissue compartment, but the annotator may set one (or even several) reference point for each tissue compartment that is identifiable in the given

tissue sample. Each tissue compartment may be associated with an index color, so that the annotation can be interpreted as an index color image that resolves the tissue structure. In case of a conflicting position where several tissue compartments match the annotation, the position may be associated with the tissue compartment whose similarity map achieves the highest intensity.

In order to validate similarity maps for image annotation, we overlaid the spectral image with a reference segmentation obtained from a supervised classifier using a well-established set of training spectra [3]. Reference points for the different tissue components were set by a human operator, aiming to reproduce the reference segmentation as good as possible. While the annotation thus achieved may not be optimal in the sense that a different choice of reference points achieve a higher accuracy, it simulates a segmentation that may realistically achieved by a histopathologist.

Hierarchical clustering

We employed two variants of hierarchical clustering. First, we hierarchically clustered spectra using Ward's approach [26] based on two different distance measures. First, we employed the well-established and widely used correlation distance (i.e., one minus correlation coefficient) and, second, we used the power metric $d_p(X, Y) = 1 - \sigma_X(Y)$ obtained using Eq. (1). As a further flavor of hierarchical clustering, we performed *hierarchical two-means*, i.e., recursively bipartitioning the dataset into two groups using two-means clustering in a top-down fashion. In each round of two-means clustering, the best subdivision among five repetitions on different random initialization was used for the next round of subdivision. For Ward's clustering, we utilized the (parallelized) implementation provided by the *Statistics toolbox* of *Matlab* version 7.11. Hierarchical two-means clustering was implemented using *k*-means clustering provided by *Matlab*.

Validation of hierarchical clustering

When performing hierarchical clustering on curated training data with training spectra for tissue components $1, \dots, K$, a dendrogram obtained from an "ideal" hierarchical clustering would contain one vertex v_i for each $i = 1, \dots, K$ such that all spectra below v_i belong to class i . In order to measure to what degree a dendrogram D obtained by Ward's clustering achieves this criterion, we identify vertices v_1, \dots, v_K in D that approach this goal as far as possible. As detailed below, this can be achieved based on ideas behind so-called *tree-assignments* recently introduced in a different context in [16,17,27].

The main idea behind validating how well a given dendrogram reflects a given reference partitioning of a set of spectra is to utilize measures for comparing partitionings, such as accuracy or the popular Rand index (RI).

Once such a measure is chosen, we determine a partitioning supported by the dendrogram that maximizes this measure. This approach is in line with the *VI-Cut* introduced in [24], which determines a partitioning that maximizes variation of information as an information theoretic measure for cluster validity. In terms of validation of infrared image segmentation and annotation, however, *VI-Cut* does not allow to control the depth of annotation. In fact, *VI-Cut* may in the end perform its validation on a segmentation derived from the dendrogram that realistically may not be recoverable by a human annotator.

Measures for comparing partitionings

As our main validity measure for comparing partitionings, we use the *Rand index* [19], which is a well-established measure to compare two partitionings in the context of cluster validation [28]. The Rand index is defined for two partitionings \mathcal{C} and \mathcal{C}' that partition the set $\{1, \dots, n\} = C_1 \cup \dots \cup C_k = C'_1 \cup \dots \cup C'_\ell$. Following the notation from [28], the *Rand index* (RI) is based on the indicator function

$$e(i, j) = \begin{cases} 1 & i, j \in C_a \text{ for some } a \in \{1, \dots, k\} \\ 0 & \text{otherwise.} \end{cases}$$

and $e'(i, j)$ correspondingly equal to one if i and j are in the same class in \mathcal{C}' and 0 otherwise. We can now further define

$$\begin{aligned} n_{11} &= |\{(i, j) \mid e(i, j) = e'(i, j) = 1\}| \\ n_{00} &= |\{(i, j) \mid e(i, j) = e'(i, j) = 0\}|, \\ n_{01} &= |\{(i, j) \mid e(i, j) = 0, e'(i, j) = 1\}| \\ n_{10} &= |\{(i, j) \mid e(i, j) = 1, e'(i, j) = 0\}|, \end{aligned}$$

which finally yields the *Rand index*

$$\mathcal{R}(\mathcal{C}, \mathcal{C}') = \frac{2(n_{11} + n_{00})}{n(n-1)}.$$

We will also utilize the Mirkin metric, which as a close relative to the Rand index is defined as

$$\mathcal{M}(\mathcal{C}, \mathcal{C}') = \sum_{i=1}^k |C_i|^2 + \sum_{j=1}^{\ell} |C'_j|^2 - 2 \sum_{i=1}^k \sum_{j=1}^{\ell} m_{ij}^2,$$

where $m_{ij} = |C_i \cap C'_j|$. Obviously, one can compute the Rand index easily from the Mirkin metric [28] using

$$\mathcal{R}(\mathcal{C}, \mathcal{C}') = 1 - \frac{\mathcal{M}(\mathcal{C}, \mathcal{C}')}{n(n-1)}. \quad (2)$$

Determining optimal partitionings supported by a dendrogram

Given a dendrogram with n leaves and a reference partitioning \mathcal{C}' that partitions the numbers $\{1, \dots, n\}$, we now aim to use the dendrogram to obtain a partitioning \mathcal{C} that maximizes the Rand index between \mathcal{C} and \mathcal{C}' . We allow to derive a partitioning from the dendrogram by assigning a

class label to vertices in the dendrogram, so that all leaves below a labelled vertex v will belong to the assigned class. To prevent assignments of leaves to more than one class, no ancestor or descendant of an assigned vertex can be further assigned to a class.

Eq. (2) shows that indeed it is sufficient to minimize the Mirkin metric rather than maximizing the Rand index. Furthermore, the Mirkin metric is composed of 3 parts. Since \mathcal{C}' is the reference partitioning, $\sum_{j=1}^{\ell} |C'_j|^2$ is constant. Thus we only need to minimize the left 2 parts:

$$\begin{aligned} \mathcal{M}' &= \sum_{i=1}^k |C_i|^2 - 2 \sum_{i=1}^k \sum_{j=1}^{\ell} m_{ij}^2 \\ &= \sum_{i=1}^k \left(|C_i|^2 - 2 \sum_{j=1}^{\ell} m_{ij}^2 \right). \end{aligned}$$

Let $w_i = |C_i|^2 - 2 \sum_{j=1}^{\ell} m_{ij}^2$. Then

$$\mathcal{M}' = \sum_{i=1}^k w_i.$$

Here w_i is the weight associated with class C_i . $|C_i|$ is the number of leaves underneath vertex v_i and m_{ij} is the number of points shared by cluster C_i and C'_j . Thus, the values w_i can be computed easily and quickly. The terminology introduced above suggests the following integer linear programming to identify an optimal partitioning:

$$\min \sum_{i=1}^p w_i X_i \tag{3}$$

$$\text{s.t. } \sum_{q=1}^k X_{v_q} = 1 \quad \text{for each root-leaf path } (v_1 \dots v_k) \tag{4}$$

$$\sum_{i=1}^p X_i = Q \tag{5}$$

$$X_i \in \{0, 1\} \quad \forall 1 \leq i \leq p \tag{6}$$

Here, p is the number of vertices in the dendrogram, w_i is the gained Mirkin metric if there is a cut at vertex v_i and X_i is a binary variable. $X_i = 1$ indicates that there is a cut at vertex v_i . Finally, Q is the parameter that controls how many vertices may be assigned overall in the partitioning, thus controlling the depth of annotation: A small value of Q means the “annotator” has to choose large high vertices in the dendrogram to obtain the partitioning, a large value of Q means that the partitions can be merged from many small segments in lower parts of the dendrogram.

Once a tree-assignment has been obtained, it is useful to obtain a partitioning of the dataset where each partition

is assigned one of the classes in the reference partitioning \mathcal{C}' . Such class assignment can be used to associate an accuracy of the segmentation \mathcal{C} , and in case of an image dataset can be used to produce an index color image. In order to obtain such class assignment, we follow a straightforward majority vote approach: Whenever a vertex v_i is active, i.e., $X_i = 1$, we need to associate the data points at the leaves below v_i with a class. By considering the labels of these q data points $x_{i,1}, \dots, x_{i,q}$ in the reference partitioning \mathcal{C}' , we determine the label which occurs most often, and associate it with all leaf data points $x_{i,1}, \dots, x_{i,q}$.

Our tree-assignment implementation is based on the *Matlab* interface to version 5.5 of *lpsolve*. In order to limit the size of the ILP and avoid assignments to very low-level vertices, only the topmost 255 vertices in each dendrogram were allowed to be assigned. Note that this cutoff is far beyond what could be utilized in an HCA based annotation by histologists, as the resulting pre-segmented index color image involves at most 128 different index colors and thus appears highly fragmented. Thus, vertices located even lower in the dendrogram can be considered as un-identifiable in practice by an annotator. Meanwhile, we only need to compute the topmost 255 vertices in hierarchical two-means, which can reduce the running time even further.

If applied to a training dataset where each spectrum is assigned with a class label, the result of the tree-assignment reads as a re-classification of the training dataset. Thus, we can apply any validation measure used for measuring the quality of supervised classifiers. In particular, we can mimic validation schemes such as Monte-Carlo-type cross validation by repeatedly subsampling from the training dataset. In Results Section, we extensively utilize this idea to validate hierarchical clustering in comparison to both supervised classifiers and similarity-map based annotation.

Datasets

For our computational studies, we utilized a colon tissue spectral dataset derived from [3]. The dataset consists of a training data set comprising 23,278 pixel spectra grouped into 14 classes of tissue components, along with three large spectral images displaying 854×502 , 576×672 and 832×416 FT-IR pixel spectra of three tissue sections. The images will henceforth be referred to as *120514*, *88180* and *colon_p53_active*, respectively. The spatial resolution is $5.5 \mu\text{m}/\text{px}$. Following common practice in infrared image analysis, spectra exhibiting a weak signal or strong noise, e.g. resulting from holes or cracks in the tissue section or other artifacts, are discarded in a preprocessing step. This affects roughly 10% of all image spectra; for image *120514*, e.g., 8.24% of the image spectra are not considered for further analysis.

Based on the training dataset, a Random Forest classifier has been trained (for details, refer to [3]), yielding a segmented version of the spectral images that assigns one of the 14 trained classes to each pixel, see Figure 3(A). The training data set contains well-curated spectra and has been validated in detail in [3] by further experimental evidence using fluorescence microscopy. Furthermore, it was shown in [3] that the segmentations obtained from this supervised classifier resolve histopathologically relevant details such as the *lamina muscularis mucosae*. Following the general difficulty to obtain ground truth for biological image data [29], we used this fluorescence-validated and histopathologically well supported segmentation as a ground truth segmentation to quantitatively compare with segmentations obtained from similarity maps and hierarchical clustering algorithms. We may consider the cross-validation accuracy of the random forest of 94.92% on the training data as an estimate on the accuracy of our reference data.

Results and discussion

To measure the performance of the classification and segmentation methods introduced here with other methods, we used the mean accuracy achieved in a Monte-Carlo type validation scheme whenever applicable.

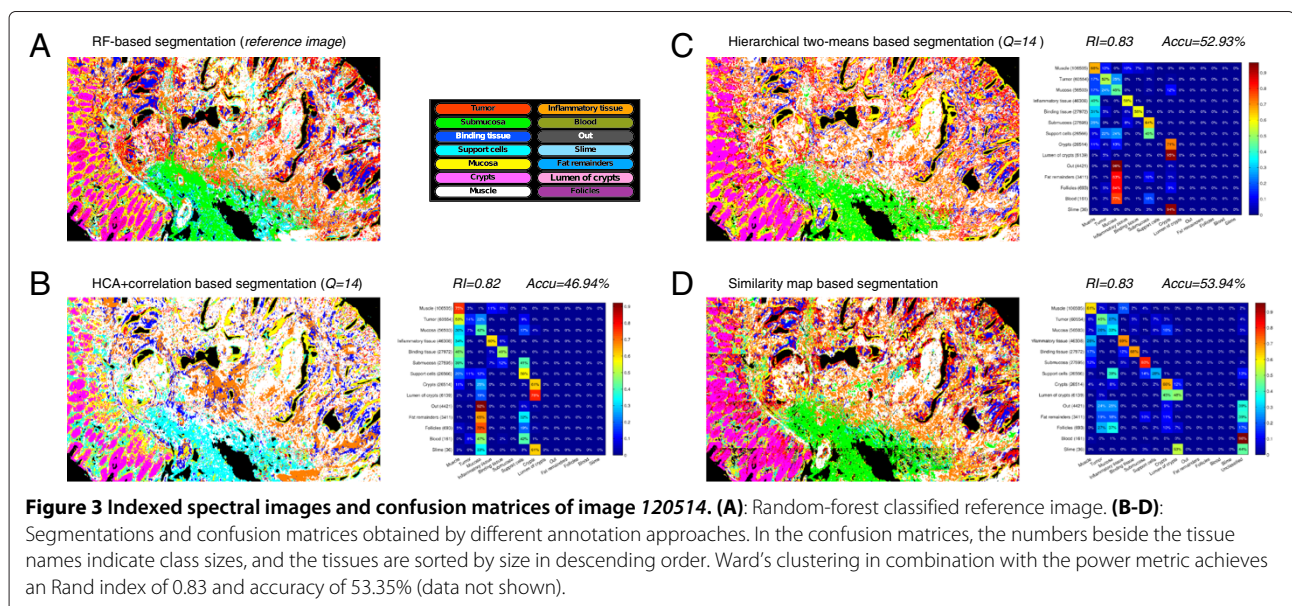
Validation of tree assignments

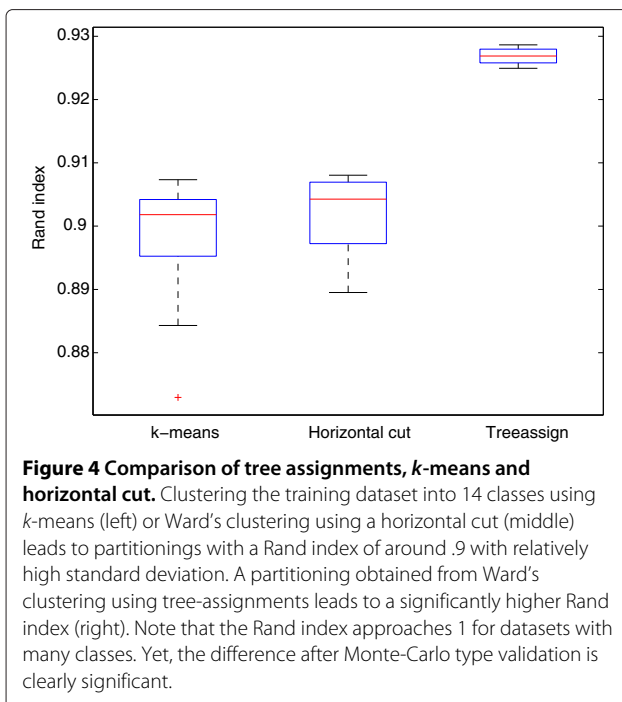
We compared segmentations obtained from tree assignments, *k*-means and horizontal cut (see Figure 4). In this case, horizontal cut performs slightly better than *k*-means. While non-horizontal cut using tree assignments gets much higher Rand index than the other two methods. Our results further confirmed the previous findings [7] in a systematic and quantitative way.

Validation of similarity maps

We applied the *Lasagne* software to all the three spectral images, using the random forest classifications based on well-curated training data as reference segmentations. These RF-based reference segmentations were visually reconstructed as good as possible using the *Lasagne* software by a human operator. We allowed the operator to specify up to two reference pixel spectra per class. In the resulting segmentation of image 120514 (Figure 3), 376,718 (95.76%) out of the 393,378 non-background pixel spectra were assigned to one of the classes in the training dataset. The smallest five classes, namely *out*, *fat remainders*, *follicles*, *blood* and *slime* could not be properly identified as either too few spectra belong to this class (36 spectra for slime) or their location patterns were spectrally not unambiguously resolved by the *Lasagne* software. Yet, the resulting segmentation assigns 53.94% out of 393,378 pixel spectra to the correct class. This accuracy is higher than the accuracy achieved by either variant of HCA, where at most 53.35% of the spectra were assigned correctly by Ward's clustering with the power metric. From the confusion matrices, we can see that both HCA based segmentations and similarity maps based segmentation perform better for big classes than for small classes. What is different is that for small classes that are difficult to identify, *Lasagne* rejects to assign any class label while HCA based methods make wrong assignments. In Figure 3(B), submucosa was totally mistakenly identified as either support cells or muscle, which is undesirable.

Figure 3 shows the RF-segmented reference image, tree assignments based segmentation and the *Lasagne*-reconstruction image for dataset 120514. Corresponding results for the other two datasets 88180 and *colon_p53_active* are shown in Additional files 6 and 7.





For dataset *88180*, the Rand index is equivalent between similarity maps and either variant of HCA (.75), while the accuracy is slightly higher for HCA based segmentations ($\geq 59.39\%$ for HCA vs. 56.23% for similarity maps). For dataset *colon_p53_active*, HCA accuracies are significantly higher ($\geq 69.21\%$ vs. 41.68%). Although HCA based segmentations received higher overall accuracies than similarity map based segmentation, many details of the tissue structure are lost. Due to the majority vote approach of class assignment subsequent to the tree-assignment based validation, they are more likely to mistakenly recognize small tissue classes as big tissue classes. This property may cause problems for samples containing unbalanced proportions of tissue classes. Furthermore, our validation is conservative in the sense that HCA is validated by a segmentation that algorithmically mimics the annotation that the “best possible annotator” could obtain from the given dendrogram, whereas the similarity map relies on a real human annotator to visually reproduce the ground truth segmentation.

Validation of different hierarchical clustering approaches

We applied and evaluated tree assignments using different depths of segmentation $Q = 14, 16, \dots, 42$ (see Figure 5 and Additional file 8). Both Rand index and accuracy increase with larger values of Q in essentially all cases. However, accuracy increases faster than the Rand index, which may be due to the relatively large number of 14 groups in our dataset, where the Rand index tends to approach 1. Hierarchical two-means performs worse than

Ward's method on training data, while comparable or even slightly better on image *120514*. In general, we may conclude that hierarchical two-means works well on image data, and using the power metric gives a slight, but not significant advantage over the established and widely used correlation distance on both image and training data. As to be expected, the accuracy achieved by unsupervised HCA using either distance measure is much smaller than the 94.92% accuracy obtained from a supervised random forest.

Beside validation measures, the running time required for obtaining clustering results is of high practical relevance. While not investigated in further detail, clustering roughly half a million image spectra using hierarchical two-means takes only few hours without parallelized computation, while Ward's clustering consumes more than one week of computation time using up to 64 CPUs in parallel.

Conclusions

We have introduced two novel concepts in the context of annotating FT-IR microspectroscopic images. First, we proposed a quantitative validation of hierarchical clustering schemes commonly employed during spectral image annotation. Second, we described and validated interactive similarity maps as an alternative to clustering-based image annotation.

Similarity maps for vibrational microscopy image segmentation

Our contribution on interactive similarity maps suggests that there are viable alternatives to this “clustering paradigm”. As our findings suggest, annotations obtained using similarity maps may achieve similarly accurate as annotations based on hierarchical clustering. Compared to the costs of computing time and memory that are still significant even for the more efficient hierarchical two-means, similarity maps require no preprocessing beyond the commonly performed low-level normalization or baseline correction. Implemented on a GPU, recomputing the similarity map after an interactive change of a reference point can be done within fractions of a second even on large ($> 500,000$ spectra) infrared images.

While visually identifying reference points is an intuitive concept addressing the histologist or pathologist not requiring any explicit computational expertise, this contribution provides a proof of concept based on quantitative validation. Establishing it to the level of a routine task for histologists or pathologists in larger scale studies is a perspective that should be encouraged by our positive quantitative validation of the approach. Both similarity-map based exploration and annotation and the concept of tree-assignments introduced here may be equally useful

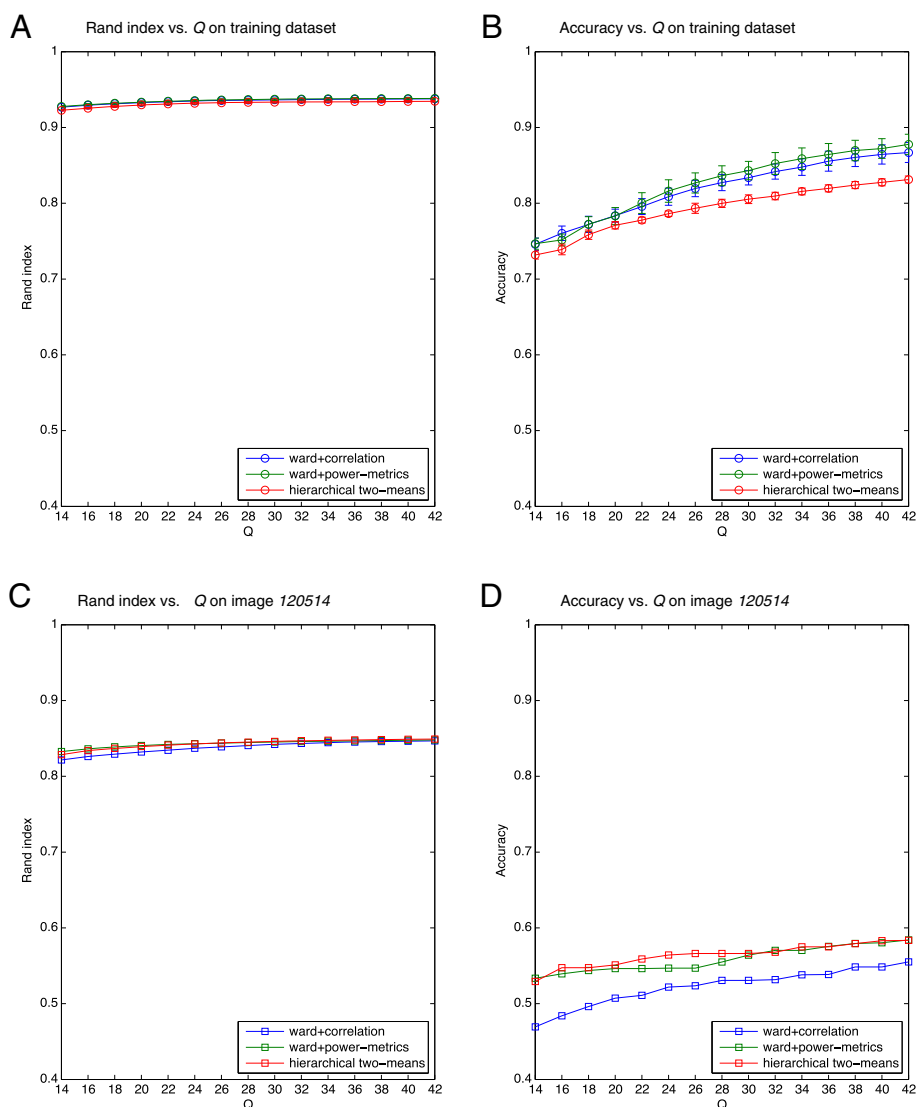


Figure 5 Comparison of different hierarchical clustering approaches under varying the depth of segmentation Q . Three hierarchical clustering schemes are evaluated in terms of Rand index and accuracy on both the training dataset (**A-B**) and image 120514 (**C-D**). 10-fold Monte-Carlo cross validation is performed on the training dataset (the error bar indicates standard deviation).

for Raman [9] and CARS [10] microscopy, which is worthwhile to explore in future contributions.

Clustering in vibrational microscopy image segmentation

Along our contribution to quantitatively compare unsupervised infrared image segmentation strategies, we have provided a validation scheme for hierarchical clustering that matches the assumptions behind spectral image annotation, which turned out to be a non-trivial task in itself. As hierarchical clustering is arguably the most commonly used basis for infrared image annotation, this contribution is particularly important for systematically quantifying performance of different methods, rather

than comparing by qualitative visual inspection. One of the immediate consequences we obtain is that the traditionally used Ward's clustering may be substituted without significant loss of quality by hierarchical two-means for image segmentation. As the latter is much more time and memory efficient, this finding will make it much more practical to work with large spectral images. Being able to handle larger numbers of spectra without compromising in terms of accuracy becomes increasingly important in multispectral microscopy. In fact, the sizes of images keep growing with new generations of FT-IR microscopes and array detectors, or when working on confocally measured stacks of Raman or CARS images.

Turning dendrograms into segmentations or partitionings

Finally, the idea of determining non-horizontal cuts in dendrograms and the cross-validation scheme based on this idea may be of further interest in infrared microscopy and beyond. Although not explored in this contribution, tree assignments also allow to compare two (or more) dendrograms by identifying an optimal set of classes supported by both dendrograms, rather than matching a fixed segmentation into one dendrogram. While this can be achieved by relatively simple modifications of the integer linear programming and the weighting scheme provided here, exploration is left for future contributions.

Additional files

- Additional file 1:** A video showing interactive annotation of a spectral image using *Lasagne* (part 1).
- Additional file 2:** A video showing interactive annotation of a spectral image using *Lasagne* (part 2).
- Additional file 3:** A video showing interactive annotation of a spectral image using *Lasagne* (part 3).
- Additional file 4:** A video showing interactive annotation of a spectral image using *Lasagne* (part 4).
- Additional file 5:** A video showing interactive annotation of a spectral image using *Lasagne* (part 5).
- Additional file 6:** Indexed spectral images and confusion matrices of image 88180.
- Additional file 7:** Indexed spectral images and confusion matrices of image *colon_p53_active*.
- Additional file 8:** Comparison of different HCA approaches on image 88180 and *colon_p53_active*.

Abbreviations

FT-IR: Fourier transform infrared; H&E: Hematoxylin and Eosin; HCA: hierarchical clustering analysis; ISMs: interactive similarity maps; RF: random forest; ILP: integer linear programming; RI: Rand index.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

QZ implemented clustering and validation algorithms, conducted computational experiments, drafted the manuscript, and participated in design of the study. CY participated in algorithm implementation. FG and AK prepared datasets. PS implemented software for interactive similarity maps. KG coordinated the study. AM conceived of the study, participated in its design, and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research was supported by the *Protein Research Unit Ruhr within Europe* (PURE) from the Ministry of Science and Technology, North Rhine-Westphalia, Germany. AM was supported by a Chinese Academy of Sciences Visiting Professorship for Senior International Scientists (grant No. 2011T1S11).

Author details

¹Department of Biophysics, CAS-MPG Partner Institute and Key Laboratory for Computational Biology, 320 Yueyang Road, 200031 Shanghai, China.

²Department of Biophysics, Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, Germany.

Received: 3 May 2013 Accepted: 7 November 2013

Published: 20 November 2013

References

1. Lasch P, Haensch W, Lewis EN, Kidder LH, Naumann D: **Characterization of colorectal adenocarcinoma sections by spatially resolved ft-ir microspectroscopy.** *Appl Spectrosc* 2002, **56**(1):1–9.
2. Steller W, Einenkel J, Horn LC, Braumann UD, Binder H, Salzer R, Krafft C: **Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging.** *Anal Bioanal Chem* 2006, **384**(1):145–154.
3. Kallenbach-Thieltges A, Großberüschkamp F, Mosig A, Diem M, Tannapfel A, Gerwert K: **Immunohistochemistry, histopathology and infrared spectral histopathology of colon cancer tissue sections.** *J Biophotonics* 2013, **6**(1):88–100.
4. Trevisan J, Angelov PP, Carmichael PL, Scott AD, Martin FL: **Extracting biological information with computational analysis of fourier-transform infrared (ftir) biospectroscopy datasets: current practices to future perspectives.** *Analyst* 2012, **137**(14):3202–3215.
5. Lasch P, Diem M, Hänsch W, Naumann D: **Artificial neural networks as supervised techniques for ft-ir microspectroscopic imaging.** *J Chemometrics* 2006, **20**(5):209–220.
6. Bird B, Miljkovic M, Romeo MJ, Smith J, Stone N, George MW, Diem M: **Infrared micro-spectral imaging: distinction of tissue types in axillary lymph node histology.** *BMC Clin Pathol* 2008, **8**(1):8.
7. Lasch P, Haensch W, Naumann D, Diem M: **Imaging of colorectal adenocarcinoma using ft-ir microspectroscopy and cluster analysis.** *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 2004, **1688**(2):176–186.
8. Kannan S, Ramathilagam S, Sathya A, Pandiyarajan R: **Effective fuzzy c-means based kernel function in segmenting medical images.** *Comput Biol Med* 2010, **40**(6):572–579.
9. Turrell G, Corset J: *Raman Microscopy: Developments and Applications.* San Diego: Academic Press; 1996.
10. Freudiger CW, Min W, Saar BG, Lu S, Holtom GR, He C, Tsai JC, Kang JX, Xie XS: **Label-free biomedical imaging with high sensitivity by stimulated raman scattering microscopy.** *Science* 2008, **322**(5909):1857–1861.
11. Matthäus C, Chernenko T, Newmark JA, Warner CM, Diem M: **Label-free detection of mitochondrial distribution in cells by nonresonant raman microspectroscopy.** *Biophys J* 2007, **93**(2):668–673.
12. Dress A, Lokot T, Schubert W, Serocka P: **Two theorems about similarity maps.** *Ann Combinatorics* 2008, **12**(3):279–290.
13. Serocka P: **Visualization of high-dimensional biomedical image data.** In *Advances in Multimedia Information Processing—PCM 2007.* Berlin Heidelberg: Springer; 2007:475–482.
14. Schubert W, Gieseler A, Krusche A, Serocka P, Hillert R: **Next-generation biomarkers based on 100-parameter functional super-resolution microscopy tis.** *New Biotechnol* 2011, **29**(5):599–610.
15. Jain AK, Dubes RC: *Algorithms for Clustering Data.* Prentice-Hall Advanced Reference Series. Upper Saddle River: Prentice Hall PTR; 1988.
16. Mosig A, Jäger S, Wang C, Nath S, Ersoy I, Palaniappan K, Chen SS, et al.: **Tracking cells in life cell imaging videos using topological alignments.** *Algorithms Mol Biol* 2009, **4**(1):10.
17. Xiao H, Li Y, Du J, Mosig A: **Ct3d: tracking microglia motility in 3d using a novel cosegmentation approach.** *Bioinformatics* 2011, **27**(4):564.
18. Halkidi M, Batistakis Y, Vazirgiannis M: **On clustering validation techniques.** *J Intell Inf Syst* 2001, **17**:107–145.
19. Rand WM: **Objective criteria for the evaluation of clustering methods.** *J Am Stat Assoc* 1971, **66**(336):846–850.
20. Hubert L, Arabie P: **Comparing partitions.** *J Classif* 1985, **2**(1):193–218.
21. Meilă M: **Comparing clusterings—an information based distance.** *J Multivariate Anal* 2007, **98**(5):873–895.
22. Friedman J, Hastie T, Tibshirani R: *The Elements of Statistical Learning, 2nd edn.* New York: Springer; 2008. Chap. 14.3.12 Hierarchical Clustering.
23. Dotan-Cohen D, Melkman AA, Kasif S: **Hierarchical tree snipping: clustering guided by prior knowledge.** *Bioinformatics* 2007, **23**(24):3335–3342.
24. Navlakha S, White J, Nagarajan N, Pop M, Kingsford C: **Finding biologically accurate clusterings in hierarchical tree decompositions using the variation of information.** *J Comput Biol* 2010, **17**(3):503–516.
25. Bruzese D, Vistocco D: **Cutting the dendrogram through permutation tests.** In *Proceedings of COMPSTAT2010:* Physica-Verlag HD; 2010:847–854. COMPSTAT 2010 Book of Abstracts, 62.

26. Ward Jr JH: **Hierarchical grouping to optimize an objective function.** *J Am Stat Assoc* 1963, **58**(301):236–244.
27. Xiao H, Zhang M, Mosig A, Leong H: **Dynamic programming algorithms for efficiently computing cosegmentations between biological images.** In *Algorithms in Bioinformatics*. Berlin Heidelberg: Springer; 2011:339–350.
28. Wagner S, Wagner D: **Comparing clusterings: an overview.** Technical Report 2006-4, Universität Karlsruhe, Fakultät für Informatik. 2007
29. Peng H, Chung P, Long F, Qu L, Jenett A, Seeds AM, Myers EW, Simpson JH: **Brainaligner: 3d registration atlases of drosophila brains.** *Nat Methods* 2011, **8**(6):493–498.

doi:10.1186/1471-2105-14-333

Cite this article as: Zhong *et al.*: Similarity maps and hierarchical clustering for annotating FT-IR spectral images. *BMC Bioinformatics* 2013 **14**:333.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

