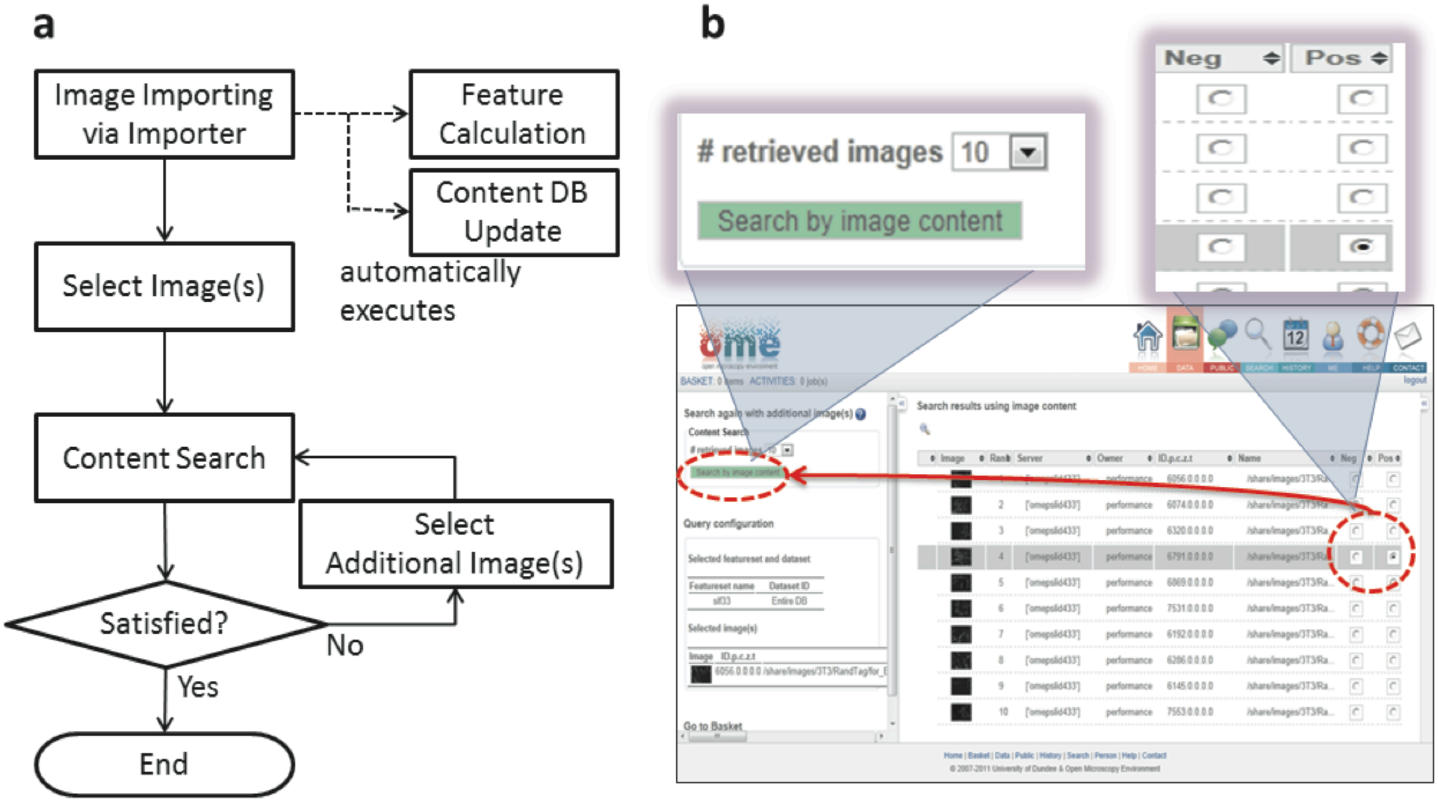
Supplementary information

OMERO.searcher: Content-based image search for microscope images

Baek Hwan Cho[1], Ivan Cao-Berg[1], Jennifer Ann Bakal[2], and Robert F. Murphy[1,2,3,4,5]

[1]Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[2]Center for Bioimage Informatics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[3]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[4]Department of Machine Learning, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[5]Freiburg Institute for Advanced Studies, Albert Ludwig University of Freiburg, Germany

Corresponding author: Robert F. Murphy (murphy@cmu.edu)



BASKET: 0 items ACTIVITIES: 0 job(s)

Search again with additional image(s)

Content Search

retrieved images 10

Search by image content

Query configuration

Selected featureset and dataset

Featureset name	Dataset ID
slf33	Entire DB

Selected image(s)

Image	ID.p.c.z.t
	6056.0.0.0.0 /share/images/3T3/RandTag/for_E

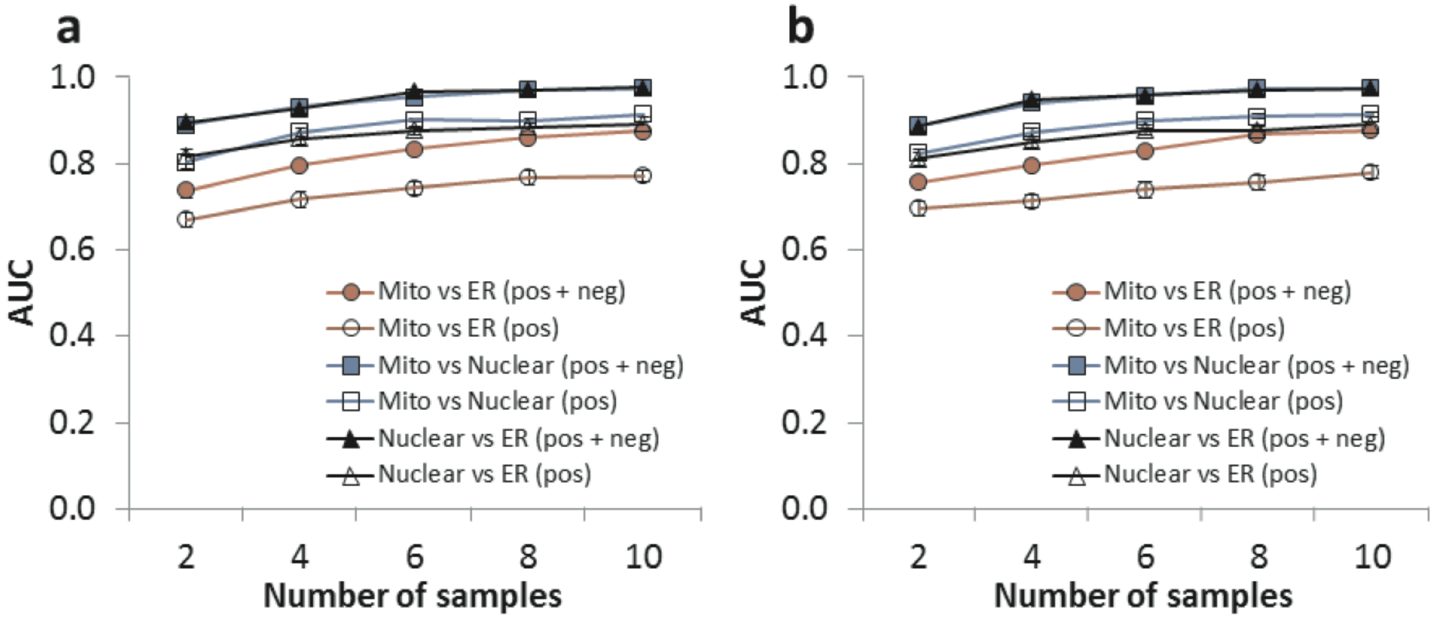
Search results using image content

	Image	Rank	Server	Owner	ID.p.c.z.t	Name	Neg	Pos
		1	['omepslid433']	performance	6056.0.0.0.0	/share/images/3T3/Ra...	○	○
		2	['omepslid433']	performance	6074.0.0.0.0	/share/images/3T3/Ra...	○	○
		3	['omepslid433']	performance	6320.0.0.0.0	/share/images/3T3/Ra...	○	○
		4	['omepslid433']	performance	6791.0.0.0.0	/share/images/3T3/Ra...	○	●
		5	['omepslid433']	performance	6869.0.0.0.0	/share/images/3T3/Ra...	○	○
		6	['omepslid433']	performance	7531.0.0.0.0	/share/images/3T3/Ra...	○	○
		7	['omepslid433']	performance	6192.0.0.0.0	/share/images/3T3/Ra...	○	○
		8	['omepslid433']	performance	6286.0.0.0.0	/share/images/3T3/Ra...	○	○
		9	['omepslid433']	performance	6145.0.0.0.0	/share/images/3T3/Ra...	○	○
		10	['omepslid433']	performance	7553.0.0.0.0	/share/images/3T3/Ra...	○	○

Home | Basket | Data | Public | History | Search | Person | Help | Contact

© 2007-2011 University of Dundee & Open Microscopy Environment

Supplementary Fig. 1. OMERO.searcher workflow. (a) Images are uploaded to the server using client software such as OMERO.importer. When an image is imported, features are automatically calculated and stored on the server. One or more images are then selected and used to find a set of similar images. (b) This set is displayed and the user can then refine the search by selecting additional examples of things that are desired (positive) or not desired (negative).



Supplementary Fig. 2. Results of retrieval performance tests for images at different resolutions. The set of 16,537 RandTag images from the PSLID repository (http://pslid.org) from Fig. 1b were downsampled by a factor of 4 to simulate having been acquired with 10x magnification instead of the original 40x (the original image resolution was 0.16125 microns/pixel). (a) The simulated 10x images were imported into a test OMERO database. Query images were drawn only from these 10x images and AUC values for retrieval from the remaining 10x images were calculated as in Fig. 1b. The average AUC for 10 images (5 positive and 5 negative) across all three combinations of patterns was 0.934. (b) The simulated 10x images and the original 40x images were imported into a new test OMERO database. Query images were drawn at random from this database, and the search was done using the lowest resolution of the query set (i.e., 10x unless all query images were 40x). AUC values were calculated as before. The average AUC for 10 images (5 positive and 5 negative) across all three combinations was 0.936.

Supplementary Methods

For Fig. 1a, a dataset from the ASCB Cell Image Library (http://www.cellimagelibrary.org/) was used. A total of 1,164 fluorescent microscope images were used, each annotated with one of 15 different cellular component terms describing the subcellular location pattern of the protein in the image. The images were grouped into 11 classes consisting of all combinations of annotation terms. For each image as a query, an image ranking was obtained for the remaining images; if the retrieved image had the same class (same combination of cellular component annotations) as the query image, it was considered to be a hit, otherwise a miss. A receiver operating characteristic curve was created for each ranking, in which the fraction of hits (true positive rate) was compared to the fraction of misses (false positive rate) as the number of ranked images was increased. The area under this curve (AUC) was averaged across all queries for a given class. For Fig. 1b, a total of 16,537 RandTag images from the PSLID repository (http://pslid.org) were used, each of which was annotated with one of three protein location pattern class labels. One or more images were chosen at random from one of the classes, the images of that class and one of the other classes were ranked by similarity, and an ROC constructed. This was repeated 500 times and the results were averaged. The average AUC across all three combinations for 10 images was 0.916. As a second test, the same approach was used except that one or more images were chosen from two different classes (one "positive" and one "negative). The average AUC for 10 images (5 positive and 5 negative) was 0.976.

All data and testing scripts used to create Fig. 1 and Supplementary Fig. 2 are available at http://murphylab.web.cmu.edu/software.

Supplementary Note

OMERO.searcher is available at http://murphylab.web.cmu.edu/software/searcher and can be used by any of three approaches.

1) The OMERO.searcher External Search page of one or more OMERO databases that allow external searches can be used to search for images similar to those on the local computer (the local images do not have to be in any OMERO database or in any special format). The database that is to be searched within must have OMERO.searcher installed and have enabled an External Search page. The only requirement on the client side is a compatible web browser. A public, externally searchable database has been created for demonstration and testing purposes using a subset of images from the RandTag project (Garcia Osuna et al, 2007). The search page is http://omepslid2.compbio.cs.cmu.edu/content_search. Additional public databases will be made available shortly.

2) OMERO.searcher Local Client can be run on a local computer to search through any OMERO database for which you have access privileges (including the public externally searchable database above).

3) OMERO.searcher Server can be installed on an OMERO database server.